University of Canterbury
Department of Physics and Astronomy

MapPHAN for Conformal Arc SABR: An Analysis of Normal Tissue Dosimetry

Nathan Henry

Submitted in partial fulfilment of the requirements for the degree of
Masters in Medical Physics at the University of Canterbury
March 2019

Abstract

Stereotactic Ablative Body Radiotherapy (SABR), when delivered to the lung, provides several unique challenges from a quality assurance (QA) standpoint. To date, there has been little research performed on the efficacy of QA at detecting significant dosimetric errors in organs at risk (OARs). In addition, the feasibility of using the MapCHECK phantom housed in the MapPHAN accessory for SABR QA is not fully appreciated. The aim of this study is to investigate the sensitivity and specificity of MapCHECK in MapPHAN to introduced errors in conformal arc SABR plans, with a specific focus on OAR dosimetry.

Multi-Leaf Collimator (MLC) class shift errors up to 2 mm, and isocentre shift errors up to 1 mm, were introduced to 26 simulated 6 MV X-ray conformal arc lung SABR plans on Raystation. 10 of these plans had errors introduced, and were delivered on a Varian Clinac iX linear accelerator to the MapCHECK phantom housed in MapPHAN. In order to commission the phantom for this project, the inherent angular dependency of the MapCHECK diodes had to be corrected to within ± 2%. Gamma analysis was used to compare the measurements on MapCHECK to the simulated Raystation plans, using 1%/1 mm, 2%/1 mm, 2%/2 mm, 3%/2 mm, 3%/3 mm and 5%/1 mm gamma criteria. Based on these results, Receiver Operating Characteristic (ROC) curves were generated to determine the sensitivity and specificity of MapCHECK in MapPHAN to the introduced errors. These ROC curves were then used to determine an optimal set of gamma criteria for use in QA.

It was found that the angular dependency of the diodes in MapPHAN can be reduced to within ± 2% for the central section of the phantom, making it viable for absolute dosimetric measurements of SABR plans. In general, the sensitivity and specificity of the phantom to introduced errors was highest when using 5%/1 mm gamma criteria, although other criteria provided superior accuracy at certain gamma thresholds, with 2%/1 mm being the best alternative criteria. The phantom was most sensitive to introduced MLC errors, and exhibited poor sensitivity and specificity to introduced isocentre shift errors. This was exacerbated when automatic shifts were introduced to correct for setup errors during gamma analysis.

In summary, MapCHECK in MapPHAN is a viable alternative to film for patient-specific QA of lung SABR plans. It provides sufficient resolution to take measurements in high dose gradient regions, and allows absolute dosimetry to be performed following angular dependency

corrections. Based on the results obtained here, a combination of 5%/1 mm with 95% threshold and 2%/1 mm with 90% threshold would provide good sensitivity and specificity for QA of conformal arc SABR plans. It is probable that placing restrictions on the shifts included in SNC Patient, or removing them entirely, would significantly improve detection of misalignment errors.

Acknowledgements

This project would not have been possible without the generosity of many people. I would firstly like to thank my supervisor Chris Thompson, who inspired this project and provided invaluable support and insight along the way. Steven Marsh and Jack Aylward also supervised, and were both excellent sources of guidance and inspiration at critical junctures.

David Lawson assisted in creating the cradle for the MapPHAN, around which this project was based. Daniel Gerhard provided statistical advice from the University of Canterbury, which focussed the project's direction. Both contributions were much appreciated.

Several of my colleagues were, wittingly or otherwise, an important part of this work, as I tapped into their wealth of medical physics knowledge. In particular, my fellow registrars were wonderful sources of insight and motivation. Last and not least, to friends and family. Thank you for keeping me sane, and I sincerely hope you aren't reading this in your spare time.

Abbreviations & Acronyms

3D-CRT - 3-Dimensional Conformal Radiotherapy

AAPM - American Association of Physicists in Medicine

AUC - Area Under Curve

BEV - beam's eye view

CHART - Continuous Hyperfractionated Accelerated Radiotherapy

CPE - Charged Particle Equilibrium

CT - Computed Tomography

CTV - Clinical Target Volume

DCAT - Dynamic Conformal Arc Therapy

DICOM - Digital Imaging and Communications in Medicine

DTA - distance to agreement

DVH - Dose Volume Histogram

EBRT - External Beam Radiation Therapy

FN - false negative

FP - false positive

GTV - Gross Tumour Volume

ICRU - International Commission on Radiation Units and Measurement

IMRT - Intensity Modulated Radiation Therapy

ITV - Internal Target Volume

linac - linear accelerator

MLC - Multi-Leaf Collimator

MU - Monitor Unit

NTCP - Normal Tissue Complications Probability

OAR - Organ At Risk

PDD - Percentage Depth Dose

PSQA - Patient-Specific Quality Assurance

PTV - Planning Target Volume

QA - Quality Assurance

ROC - Receiver Operating Characteristic

SABR - Stereotactic Ablative Body Radiotherapy

SBRT - Stereotactic Body Radiotherapy

SD - standard deviation

SN - sensitivity

SP - specificity

SNC - Sun Nuclear Corporation

TCP - Tumour Control Probability

TG - Task Group

TI - Therapeutic Index

TN - true negative

TP - true positive

TPS - Treatment Planning System

VMAT - Volumetric Modulated Arc Therapy

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Lung cancer

Lung cancer is the most prevalent form of malignant neoplasm in the world, and the leading cancer-related cause of mortality worldwide [4, 5]. In New Zealand, lung cancer is the leading cause of cancer-related mortality [6]; global statistics indicate that amongst females, New Zealand has one of the highest rates of lung cancer worldwide [7]. In addition, the lung is one of the most common sites for distant metastasis in many tumours. It has been estimated that almost one in three cancer patients will develop pulmonary metastases during the course of their disease [8]. From geographic and temporal analysis of lung cancer incidence worldwide, it is widely recognized that tobacco smoking is the primary carcinogenic cause of lung cancer [4], with an estimated 85% of lung cancer-related deaths caused by smoking [9]. It is likely that lung cancer will continue to be a significant public health issue for decades to come [10].

1.2 Radiotherapy

Radiotherapy is one of the primary forms of treatment for lung cancer, along with surgery, chemotherapy, and immunotherapy. It is defined as the medical delivery of ionising radiation to a target volume of tissue, such as a cancerous lesion, with the aim of damaging or destroying said

tissue [11]. External beam radiation therapy (EBRT) is a form of radiotherapy in which the radiation comes from a source external to the patient. Radiotherapy treatment plans are generated on a Treatment Planning System (TPS), which is software designed for the planning and simulation of these treatments.

1.2.1 Volume definitions

The International Commission on Radiation Units and Measurements (ICRU) has defined volumes to standardize the delineation of different volumes of tissue in radiotherapy, in the recommendation documents ICRU 50, 62 and 83 [1]. The Gross Tumour Volume (GTV) is used to define the gross demonstrable extent and location of the tumour volume, as determined by the clinician. The Clinical Target Volume (CTV) includes the GTV, along with all regions surrounding the GTV that have a particular probability of containing unobservable cancerous tissue. An additional volume needs to be considered for lung tumour treatments, since tumours typically follow a motion path that corresponds to the patient's breathing. The Internal Target Volume (ITV) is defined as the volume that encompasses the CTV, accounting for fluctuations in the position, size and shape of the volume [12]. The Planning Target Volume (PTV), in the context of lung tumour therapy, is a geometrical concept introduced for the purposes of safe treatment planning. It is a volume contoured in such a way as to shape dose distributions to ensure that the entire ITV (and by extension, CTV) receives a dose above or equal to the prescribed dose. In general, a margin of error on the order of 5 millimetres is added to the ITV in all directions in order to account for errors that may be introduced into the treatment process [12].

Any volumes of healthy tissue adjacent to the PTV are denoted as organs at risk (OARs). The aim of a radiotherapy treatment is to deliver sufficient radiation dose to the target volume, while minimizing the dose delivered to surrounding OARs as much as possible. To achieve this, the oncologist will prescribe a certain dose level (measured in J kg^{-1}, or Gray [Gy]) to be received by the PTV. They will also recommend dose limits for the OARs.

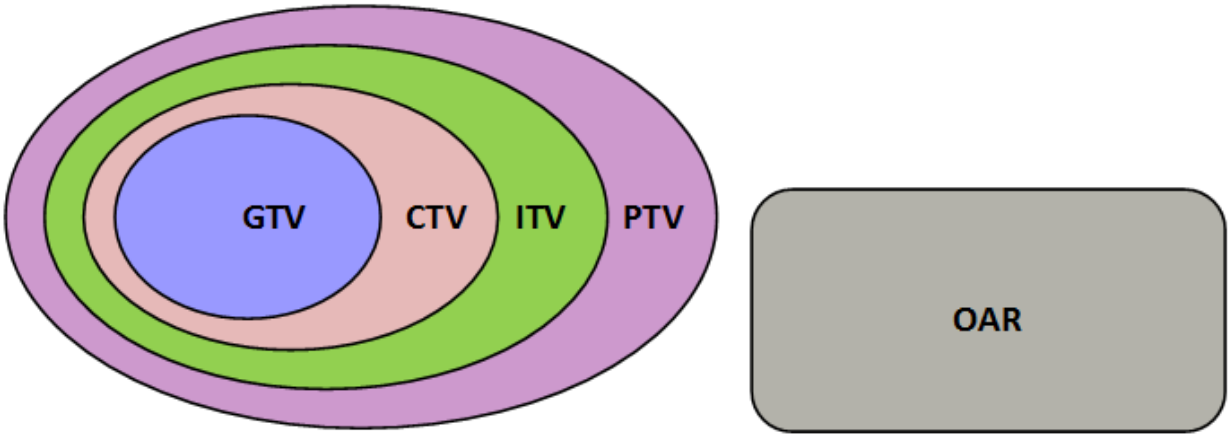


Figure 1.1: Planning volumes as denoted in the ICRU documents [1].

1.2.2 Linear accelerators

EBRT is typically delivered using a linear accelerator, or linac - a type of compact particle accelerator, which uses a high voltage waveguide to accelerate electrons onto a tungsten target; the resultant collisions with the target generates X-rays primarily by bremmstrahlung radiation [13].

Figure 1.2 indicates the main structures of which a linac is comprised. Electrons are produced via thermionic emission from the cathode of the electron gun, which is supplied by pulses of power from the pulsed modulator. At the same time, the modulator sends pulses on the order of 50 kV to the radiofrequency power source, which is typically a magnetron (or in the case of high-power linacs, a klystron), allowing the electrons to be accelerated by the waveguide [13]. The accelerating waveguide has to be evacuated of air via vacuum pump, in order to increase the mean free path of electrons within the waveguide to the point that the total electron path length is negligible by comparison. Natural divergence of these electrons due to the radial component of the electric fields in the waveguide is refocussed via the use of focusing coils in the waveguide, which produce a coaxial magnetic field to generate a pencil beam of electrons. This is guided into the head of the linac using bending magnets, typically configured to send the beam through a 270^{o} arc before striking a target made of tungsten. The interaction of the electron beam with the target produces bremsstrahlung radiation and characteristic X-rays in the mega-voltage (MV) range of energies [13].

The structures housed within the linac head are attached to a gantry, which rotates around the couch on which the patient lies, allowing for the delivery of radiation to the patient from multiple

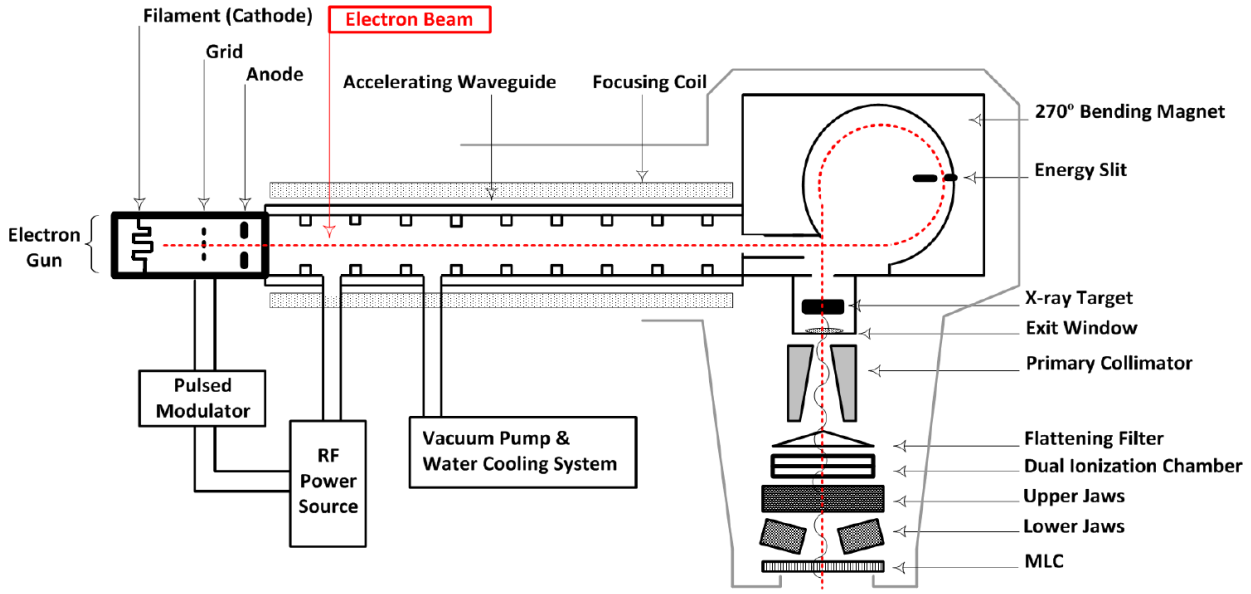


Figure 1.2: Schematic diagram of linear accelerator with waveguide mounted parallel to gantry axis of rotation (not to scale). Taken from Rowshanfarzad [2].

angles. Hence positioning the target volume at the centre of the axis of rotation of the linac will result in a high dose of radiotherapy being delivered to the tumour volume, and simultaneously minimise the dose being delivered to adjacent healthy tissue. (It is common for gantry angle to be denoted in shorthand. For instance, G230 means that the gantry is positioned at 230°.)

The linac head contains various structures that conform the beam to the target volume, allowing for further reduction of dose to OAR's. The primary collimator consists of a circular opening machined into a tungsten block, defining a maximum circular field; this can then be further confined via use of adjustable secondary collimators, consisting of two upper and two lower independent jaws which truncate the beam into a rectangular field. Between the primary and secondary collimators lies a flattening filter, which can be used to modify the beam output into a flat beam profile when measured at the level of the patient.

Further beam collimation is achieved via use of a Multi-Leaf Collimator (MLC). In the case of the Varian Clinac iX linear accelerator (Varian Medical Systems Inc., CA, U.S.A.), this is a series of 120 opposing tungsten leaves that sits behind the secondary collimating jaws of the linac. While the X- and Y-jaws serve to define the overall field size of the beam, the MLC provides higher-resolution collimation to conform the beam to the contours of the target volume. In treatments that utilise arc therapy, the MLC's leaves move in real-time to modulate the field shape as the gantry of the linac

rotates. The position of the leaves at any point in the treatment is defined by a finite number of control points, at a predefined interval - typically between 2-5°. An MLC-defined field shape is optimised for each control point, with the MLC leaf position between each control point determined via interpolation [14].

1.2.3 Interaction mechanisms of ionising radiation

Due to the high energies utilised to generate electrons in linacs - typically on the order of several MeV - modern linacs primarily produce X-rays by means of bremmstrahlung radiation. This phenomenon occurs when charged particles are accelerated by the electric field of an atomic nucleus. The kinetic energy lost by the charged particle is converted to a photon, which is emitted [15]. In the linac, the X-rays are generated in the tungsten target. The spatial distribution of X-rays around the target is such that the vast majority of photons are emitted in the same direction as the original electron beam [16].

The radiation dose absorbed by a volume of mass m is found by the relationship

$$D = \delta\bar{\varepsilon}/\delta m,$$

where $\bar{\varepsilon}$ is the sum of radiation energy entering the volume minus the energy leaving the volume, measured in Gy [16].

The three major interaction processes that govern the dose delivered by megavoltage X-ray beams are the photoelectric effect, Compton scattering, and pair production [16]. For X-ray beams generated by 6 MeV electrons, the Compton effect will be dominant for most tissue densities, although pair production becomes predominant at higher atomic numbers, as can be seen in Figure 1.3 [3].

1.2.4 Percentage depth dose curves

The radiation dose deposited as a function of depth in a medium can be graphically represented in the form of a percentage depth dose (PDD) curve. An example of a PDD is shown in Figure 1.4.

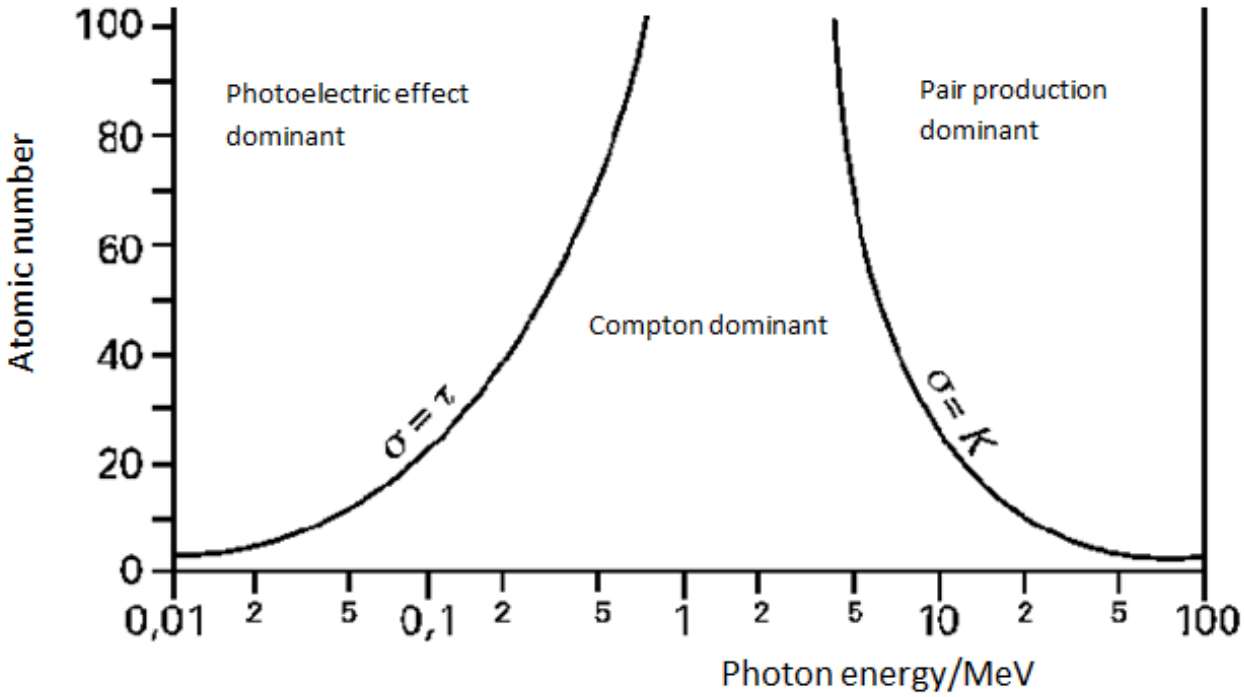


Figure 1.3: Regions of relative predominance for photoelectric effect, Compton effect, and pair production in matter. The curves represent the lines of equality for the adjacent effect regions. Modified from [3].

Several key regions can be observed in this figure. The region between $z = 0$ and $z = D_{max}$ (a depth known as z_{max}) is known as the buildup region. The relative decrease in dose for this region is due to the lack of charged particle equilibrium (CPE) at the surface of the patient. In this region, there are fewer secondary charged particles due to the absence of medium above the surface [17]. Hence at this point, the electron fluence is less than the equilibrium fluence of secondary charged particles established in the surrounding medium. CPE, where the number of electrons entering a volume is the same as that exiting the volume, is established at z_{max}, where z is approximately equal to the mean range of the secondary charged particles, and the condition of fluence equilibrium is attained. Beyond Z_{max}, transcient CPE occurs due to photon attenuation and scattering in the medium [17]. The rate of dose decay in the medium then becomes approximately proportional to the inverse square law as the distance from the source increases, while also being a function of the attenuation constants of the medium placed in the beam path [18].

When performing dosimetry, it is recommended to avoid measuring dose in the buildup region of the PDD, for several reasons. Electron contamination from the linac head is typically limited to shallow depths below z_{max}, due to the high rate of attenuation of low energy electrons in water. This,

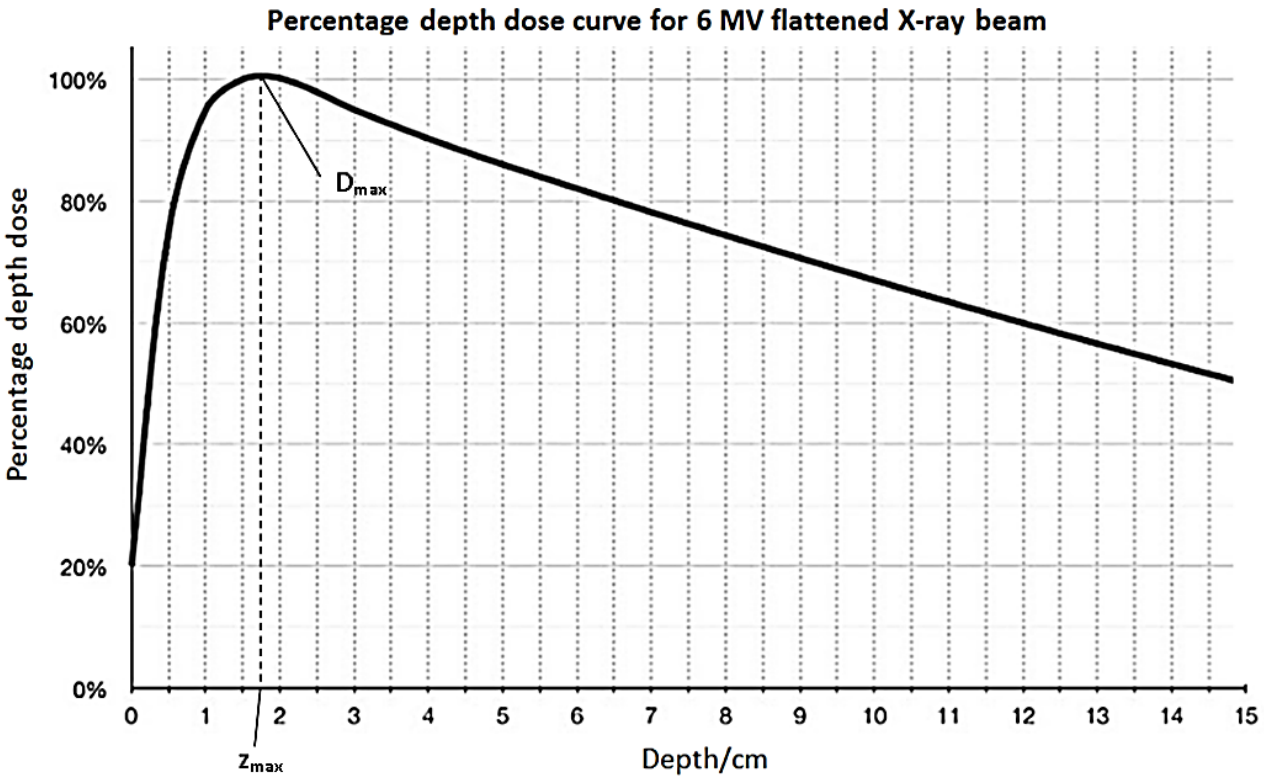


Figure 1.4: PDD for a 6 MV flattened X-ray beam in water, showing features such as the buildup region prior to D_{max} at an approximate depth of 1.5 cm, and exponential falloff at depth.

coupled with the lack of CPE in the region, increases the uncertainty in the electron dose for the buildup region. Placing the dosimeter at depths beyond z_{max} negates this issue, although the use of an extrapolation chamber is an alternative solution [19, 20]. The relatively high dose gradient in this region is another factor, since this increases the potential for geometric setup inaccuracies to increase the uncertainty in measured dose [20].

1.3 Radiobiology

Radiobiology is the study of the effects of ionizing radiation on cells. It covers a broad range of topics, from the effects of different types of radiation on cellular biology, to the effects of different types of DNA damage over varying timescales.

For every Gray of absorbed radiation dose, a cell of diameter 10 μm will experience at least 10^5 ionization events within its volume [21]. These ionization events generate free radicals, which are highly reactive and cause substantive chemical damage to the cell. Some of these lesions are

generated in DNA; most of these are then repaired by the cell. The DNA breakages that fail to repair may eventuate in cell death. While the chemical reactions that generate these DNA breakages are typically complete within approximately 1 ms of radiation exposure, the effects of the damage may be observed up to several years after the exposure [21].

1.3.1 Cell survival curves

The rate of cell survival in an irradiated volume of tissue can be expressed with some accuracy using the linear-quadratic model, which assumes that there are two components to cell kill via radiation - α, which denotes the component of cell death that is proportional to dose, and β, which denotes the component that is proportional to the square of the dose. The cell survival fraction (S) for a delivered dose (D) is given by:

$$S = \exp^{-\alpha D - \beta D^2}$$

where α and β are constants with units of Gy^{-1} and Gy^{-2} respectively [22]. It is common practice to define a constant called the α/β ratio, which defines the dose in Gy at which the two components contribute equally to cell kill [23].

α and β have been experimentally determined to varying levels in the vast majority of human tissues. By comparing the α/β ratio for tumour tissue to that for healthy tissues which also receive dose during treatment, one can determine what fractionation scheme will produce the most beneficial treatment outcome for that particular tumour-tissue combination [24]. For instance, tissues can generally be classified as either 'early responding', in that they are damaged quickly by radiotherapy but can be spared by prolonging of treatment; or 'late responding', in that they respond more slowly but at a more linearly proportional rate to radiation dose [23]. More simply put, prolonging the treatment time of radiotherapy has the greatest sparing effect on early responding tissues.

Figure 1.5 shows a typical survival curve for early and late responding tissues. As can be seen in the equation above, α comprises the linear portion of the survival curve, whereas β comprises the quadratic portion, which gives the function some curvature. Hence it can be deduced that early responding tissues have a high α/β ratio, while late responding tissues have a low α/β ratio [24].

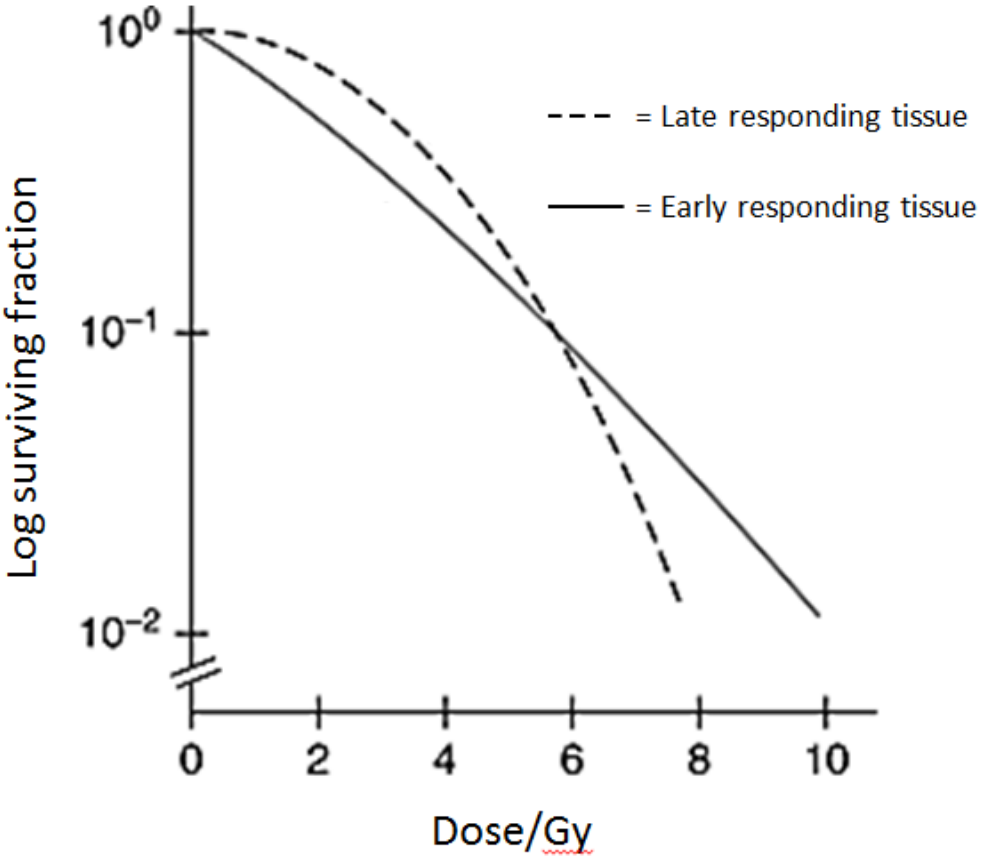


Figure 1.5: Example of surviving fraction for early and late responding tissues. Early responding tissues have a high α/β ratio, while late responding tissues have a low α/β ratio.

1.3.2 Fractionation

Since ionising radiation is absorbed by both the target tissue and healthy tissue during radiotherapy, it is necessary to plan the treatment in such a way as to minimise the damage to healthy tissue while maximising the volume of cancerous tissue that is destroyed. The therapeutic success of such a plan depends on the estimated balance of tumour control probability (TCP) and normal tissue complication probability (NTCP) that arises from the dose distribution within the patient. It also depends on the schedule with which the dose is delivered over time, also known as fractionation [25, 26].

Fractionation is necessary because current techniques are unable to generate a dose distribution for a single fraction that is sufficiently conformal for sparing of healthy tissues surrounding the target volume, while delivering a sufficiently ablative dose to the CTV. The division of a single ablative dose into multiple fractions of partially ablative doses spread out over a specified period of days

allows healthy tissue to repair in between fractions. However, this scheme also allows cancerous tissue time to heal in between fractions, meaning that a higher overall dose must be delivered to the tumour to provide an adequate rate of cell kill. The downside of this is an increased risk of secondary tumour generation, as radiation (while useful for the immediate destruction of tissue) is also a carcinogen, and an increase in radiation dose is associated with an increased risk of tumour generation later in life [27].

Hypofractionation

A fractionation scheme with dose greater than 2 Gy per fraction (considered to be the conventional scheme) is termed hypofractionation [28]. However, the advancement of radiotherapeutic techniques, equipment and treatment planning systems has significantly increased the conformality that can be achieved with a planned dose distribution. Innovations such as the introduction of Intensity Modulated Radiation Therapy (IMRT), which brought with it significant improvements in the accuracy and manoeuvrability of MLC leaves, have mitigated the risk of excessive toxicity in normal tissues, bringing down the NTCP for these tissues while simultaneously enhancing the TCP value for the tumour [28, 29]. Additionally, trials such as CHART (Continuous Hyperfractionated Accelerated Radiation Therapy) have shown that decreasing overall treatment time has significant benefit in certain cases [30, 31]. A typical hypofractionated scheme used for the treatment of non-small cell lung cancers, for example, would be 48 Gy delivered to the volume in 4 fractions, or 12 Gy per fraction.

Hypofractionation is not an optimal treatment solution for all cancers. For example, the α/β ratio is higher in brain tumours than in brain tissue, meaning that hyperfractionation is required to maximize the therapeutic ratio for radiotherapy of the brain. However, for cancers such as those found in the prostate and lung, the reverse is true: the α/β ratio is lower in the tumour than in the surrounding tissue. This allows clinicians to utilize hypofractionation, with the additional benefit of the tumour being more sensitive to fraction size. The main risk of hypofractionation is the excessive exposure of adjacent normal tissues to undesirable levels of toxicity, without enough time for sublethal repair to occur [32]. For instance, a common side effect of delivering high doses of radiotherapy to peripheral lung tumours is rib fracture, the risk of which is associated with tumour location, volume and proximity. In addition, dose metrics such as V_{160}, the volume of tumour receiving greater than 160 Gy, and $D_{4.6cc}$, the dose received by at least 4.6 cm^3 of the rib, can both

be used to predict increased risk of fracture [33].

1.4 Stereotactic Ablative Body Radiotherapy

There are several distinct radiotherapy techniques that may be utilised for the treatment of lung cancer. The technique being studied in this thesis is known as Stereotactic Ablative Body Radiotherapy, or SABR (also known as Stereotactic Body Radiotherapy, or SBRT). SABR involves the precisely targeted delivery of fewer fractions with a higher dose per fraction, utilising the α/β ratios of certain types of tumour in order to maximise TCP and NTCP outcomes for a treatment [29]. In terms of patient outcomes, SABR is considered to be one of the most effective medical techniques available for treatment of non-small cell lung cancer [34–36].

Hypofractionation in SABR has been made possible in recent years by significant improvements in kV and MV imaging technologies, allowing sub-millimetre setup accuracies to be achieved, with an overall radiation delivery precision of 1-2 millimetres [37]. SABR utilises multiple coplanar and non-coplanar beams guided by a rigid co-ordinate system with structures in place to immobilize the patient, typically using image-guided alignment during treatment to enhance accuracy. Tumour motion in the lung is normally compensated for with various immobilization, tracking, and gating techniques, although this project focusses on the accuracy of radiotherapy delivered to a static volume.

In order to treat deep-seated tumours surrounded by healthy tissue using EBRT, some ionising radiation will need to pass through healthy tissue, depositing dose along the way. The use of higher doses per fraction in SABR means that to deliver SABR safely using X-rays, one must plan the treatment in such a way as to spread the necessary dose to OARs over a greater volume of healthy tissue. Hence, one must use either multiple fields (e.g., 10 or more) or large angle arc rotations with small aperture fields, in order to limit exposure to normal tissue [29].

Non-small cell lung cancer is an example of a tumour in which the α/β ratios for the tumour and surrounding normal tissues are suitable for hypofractionation [38, 39]. Tumours in the lung are surrounded by three critical organs - namely the heart, lung and spinal cord - and the management of dose to these tissues is second only to PTV coverage in terms of importance in the treatment

planning process. Treating with SABR provides a conformal dose distribution with sufficient accuracy to generate an ablative dose to small tumours while sparing significant lung volume from excess radiation. The high dose gradients provided by SABR are especially useful for the treatment of central tumours close to the heart and spine, where the separation between the treatment volume and critical structures is too small for conventional radiotherapy to be successful [38, 40].

1.5 Treatment planning

The introduction of the computerised TPS led to rapid innovations in treatment planning. The TPS stores beam model data for simulating dose distributions in media, based on input parameters set by the user. During the commissioning process for a linac, this beam model data is calculated and measured for every combination of beam type and beam energy that the linac is capable of producing. As part of the beam model data, the standard calibration factor of the linac for each beam energy is calculated with units of cGy per monitor unit (MU), allowing the dose for each patient's treatment plan to be calculated as a certain number of MU's [16]. Using this value, the beam output of the linac may then be monitored by the ionization chambers located in the linac head, as an additional safety measure. The chamber's electrometer circuitry is typically adjusted so that 1 MU is equivalent to a dose of 1 cGy delivered to a water phantom at the depth of maximum dose on the central beam axis, using a 10 x 10 cm^2 field at a source-to-surface distance of 100 cm [16].

The most common treatment planning method still in use today is 3D conformal radiotherapy, or 3D-CRT. To optimise a 3D-CRT treatment, a planner will manually select appropriate beam angles and weightings to create a conformal dose distribution, in a process known as forward planning. This is essentially a trial-and-error approach to treatment planning [41]. In recent years, planning has evolved to the extent that optimization via inverse planning is now possible, where dose constraints are set initially and a computer optimizes the plan within human-set boundaries, via Pareto-frontier optimization [41, 42]. SABR itself can be planned with 3D-CRT, but may also be planned with one of the methods detailed in the remainder of this Section, although each method has its own set of strengths and weaknesses.

1.5.1 Treatment setup parameters

Isocentre

The fixed point around which a body rotates in 3-dimensional space is known as the isocentre of rotation for that body. The isocentre definition most relevant to treatment delivery is the radiation isocentre, which has been defined according to Zhang *et al.* as follows. At any gantry angle, a straight line between the X-ray source and centre of beam collimation (both functions of gantry angle) can be drawn. The centre of the smallest sphere that is intersected by all of these lines is the radiation isocentre [43]. This isocentre is separate from the mechanical isocentre, which has been defined as "the intersection point of the axis of rotation of the collimator and the axis of rotation of the gantry" [43]. During the installation of a linac, lasers are set up in the treatment room in such a way that their point of intersection is aligned with the machine isocentre. This intersection point is then used as the primary reference point for patient localisation during treatment, with all couch shifts performed relative to this point. The mechanical isocentre will never be perfectly aligned with the radiation isocentre, due to various factors such as gantry sag (the effect of gravity on gantry position), which can introduce errors on the order of a millimetre to the isocentric position [44].

1.5.2 Intensity Modulated Radiation Therapy

Intensity Modulated Radiation Therapy, or IMRT, is an inverse-planned radiotherapy technique. To generate an IMRT plan, a TPS will use simulated annealing to find an optimal solution for field apertures, beam directions, beam weights and MLC leaf positions, in order to achieve constraints set by the planner [41]. Human guidance is necessary to set these constraints, so that the number of Pareto-optimal solutions can be reduced in the interest of efficiency. The conformality that can be achieved with multiple IMRT fields is superior to that attainable with 3D-CRT, with a significant reduction in OAR doses made possible [41, 45].

Volumetric Modulated Arc Therapy

Volumetric Modulated Arc Therapy (VMAT) is a type of IMRT, in which the gantry rotates continuously during treatment delivery. It has a greater number of degrees of freedom than standard

IMRT, as it permits the simultaneous variation of gantry rotation speed, MLC leaf position and dose rate, while also reducing the number of MUs per fraction [14]. The plan is delivered in full 360° arcs, which are divided into control points.

A comparison of treatment techniques indicates that IMRT and VMAT typically produce a substantial reduction in the dose to OARs. [46, 47]. Although VMAT can increase the total dose received by normal tissues, it does have the advantage of spreading this dose over a greater volume of tissue, which reduces the maximum dose (and thus the NTCP) for certain OARs [14].

1.5.3 Dynamic Conformal Arc Therapy

Dynamic Conformal Arc Therapy (DCAT) is similar to VMAT, in that the gantry rotates continuously during treatment delivery. However, only MLC positions are modulated in DCAT, and the dose rate remains continuous during the treatment. The MLC shape is defined by the shape of the target volume, as seen through the beam's eye view of the source (i.e. the source as viewed from the virtual source of the linac, in the gantry head). Because of this, the MLC field size is generally greater for DCAT treatments compared to VMAT. This can lead to less uniform PTV doses with greater dose maximums than VMAT treatments, which is normally acceptable due to the high ablative doses used in SABR [48].

VMAT is generally considered to be a more precise technique than DCAT, able to generate more conformal dose distributions with lower OAR doses and more uniform PTV doses [49]. However, since VMAT treatment plans uses MLC leaves to reduce OAR doses, the control points for these plans regularly contain leaves that block the PTV. This increases the uncertainty in PTV dose compared to conformal arc plans, due to factors such as the interplay effect, where the out-of-phase interactions between MLC motion and target motion generates unpredictable differences between planned and delivered doses [50]. This is especially an issue for lung treatments, where breathing motion is significant compared to PTV size.

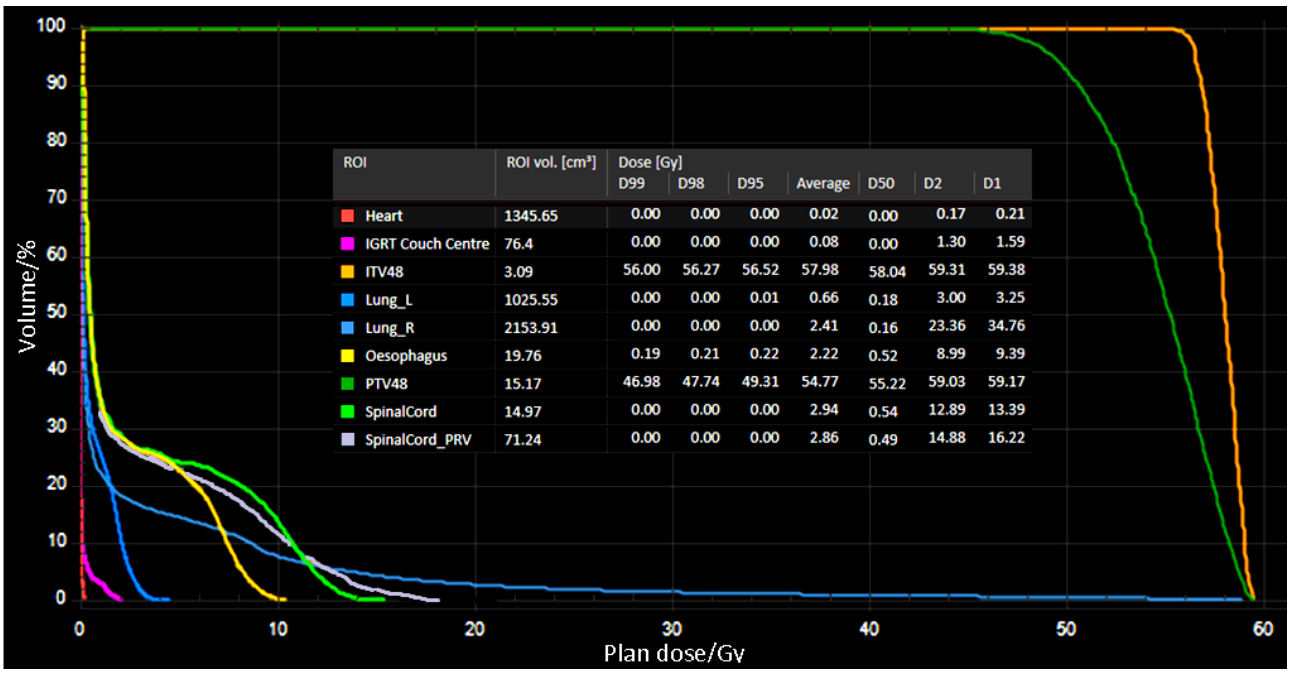
1.5.4 Dose Volume Histograms

The Dose Volume Histogram (DVH) is a tool used by planners for volume-specific analysis of a plan's dosimetry. It summarizes the dose distribution information over a 3D matrix of points in the patient's anatomy. The most commonly used version is the cumulative DVH, an example of which is shown in Figure 1.6, although there are alternatives to this, such as the differential DVH [25, 51]. The main disadvantage of the DVH is the lack of spatial information provided by the graph, which summarizes doses to each volume as a whole and hence does not provide any further information about the dose at different points in the volume [51]. Hence the DVH cannot be used as a standalone tool for plan analysis.

Figure 1.6 shows an example of a cumulative DVH for a lung SABR patient, over the entire course of treatment. Several key metrics can be derived from this figure; for instance, it can be seen that the maximum point dose being received by the PTV is approximately 60 Gy, while 95% of the PTV volume is receiving a minimum dose of 49.31 Gy. These metrics can be used to define some tolerance thresholds for the plan. In this case, the clinician may want to ensure that the spinal cord receives no more than 18 Gy, since doses above 18 Gy have been associated with increased risk of neurological deficits such as myelopathy [52]. Hence a limit may be placed on the maximum point dose that can be received by the volume 'SpinalCord_PRV', which is the Planning Risk Volume for the spinal cord (a volume containing the spinal cord contour, with an isometric 3 mm risk margin added on all sides of the volume for additional safety). This constraint can then be used to optimize the plan in such a way as to prevent the spinal cord dose from exceeding the tolerance dose.

1.6 Patient specific quality assurance

Quality assurance (QA) is an essential process in radiotherapy. One of the main roles of the medical physicist is ensuring that treatment plans are being delivered to the highest possible standard of conformality and quality. Patient-specific quality assurance (PSQA) is a branch of QA that involves simulating the delivery of a treatment plan on the machine itself, then comparing the outcomes of this treatment delivery to the predicted outcomes of the treatment plan. This is done to ensure the treatment being delivered is consistent with the prescribed plan. The measurement of expected dose



ROI	ROI vol. [cm³]	Dose [Gy] D99	D98	D95	Average	D50	D2	D1
Heart	1345.65	0.00	0.00	0.00	0.02	0.00	0.17	0.21
IGRT Couch Centre	76.4	0.00	0.00	0.00	0.08	0.00	1.30	1.59
ITV48	3.09	56.00	56.27	56.52	57.98	58.04	59.31	59.38
Lung_L	1025.55	0.00	0.00	0.01	0.66	0.18	3.00	3.25
Lung_R	2153.91	0.00	0.00	0.00	2.41	0.16	23.36	34.76
Oesophagus	19.76	0.19	0.21	0.22	2.22	0.52	8.99	9.39
PTV48	15.17	46.98	47.74	49.31	54.77	55.22	59.03	59.17
SpinalCord	14.97	0.00	0.00	0.00	2.94	0.54	12.89	13.39
SpinalCord_PRV	71.24	0.00	0.00	0.00	2.86	0.49	14.88	16.22

Figure 1.6: Example of a DVH for one of the lung SABR patients in this study, with overlaid dose metrics. This DVH was generated in the Raystation TPS.

can be performed in several ways, some of which are discussed below.

1.6.1 Ionization chambers

The ionization chamber is the most commonly used dosimeter in radiotherapy. It typically consists of an unsealed cavity of air encapsulated by a conducting electrode, with a collecting electrode at its centre. Ionizing radiation passing through this cavity ionizes air within the chamber volume and produces pairs of charged particles which are collected by the electrodes, producing a current which is measured by a connected electrometer. The number of primary ions collected is proportional to the energy deposited by the charged particle tracks in the detector volume, which can then be used to calculate the dose delivered to the chamber. An example of a CC04 ionization chamber (IBA Dosimetry, Louvain-La-Neuve, Belgium) used in this thesis is shown in Figure 1.7. The outer and inner electrodes of this chamber consist of C-552 plastic, and house an active air volume of 0.04 cm^3, a volume small enough to be approximated as a point for the purpose of point dose measurements.

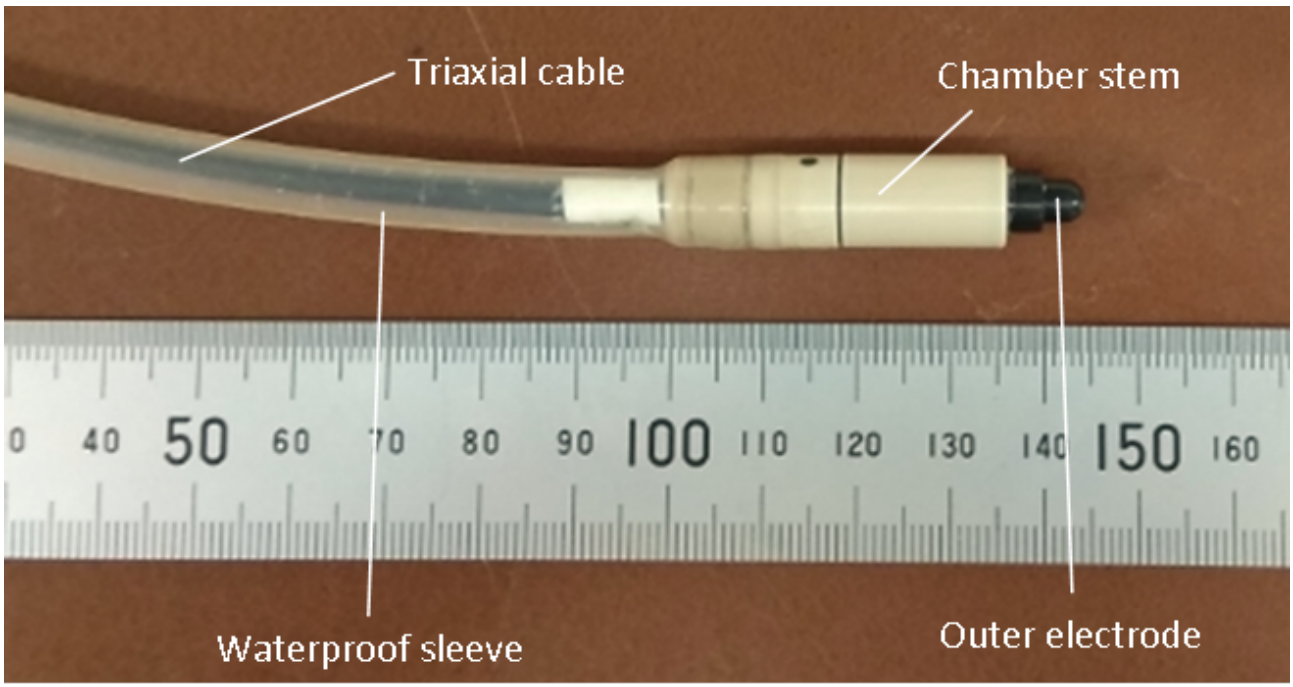


Figure 1.7: Image of CC04 ionization chamber used during this thesis.

Radiochromic film

Radiochromic film is a near tissue-equivalent dosimeter that can be used to measure a 2D dose distribution with extremely high resolution. It is approximately energy independent above 0.1 MeV, making it ideal for measurement of high energy X-ray dose distributions. Absolute dosimetry is possible following calibration of a batch of film, whereby multiple film pieces are irradiated at different levels and scanned into a computer. The development of each piece of film can be used to generate a calibration curve of film density (logarithm of opacity) versus the logarithm of exposure, from which absolute dose can be calculated. The dose distribution in the film can then be compared to the original calculated dose distribution [53].

Silicon diodes

With their small volume and high atomic number materials, diode detectors offer the advantage of higher spatial resolution and sensitivity than ionization chambers. The electron density of silicon in diodes is approximately 18,000 times greater than that of air, meaning that a silicon diode is vastly more sensitive to radiation dose than an air-cavity ion chamber, even when the diode volume is significantly lower. Diodes require regular recalibration in order to be used for absolute dosimetry, as

they acquire cumulative radiation damage over time, reducing their sensitivity by displacing Si atoms from their lattice positions.

The diode array used in this work was housed within the MapCHECK 2 phantom, model 1177 (Sun Nuclear Corporation, Melbourne, Florida), which consists of 1527 N-type SunPoint diode detectors that are uniformly spaced in a 32.0 x 26.0 cm^2 2D array, housed in acrylic casing. The shape of the array can be seen in Figure 1.8. Each detector has an active volume of 0.019 mm^3. The diodes are soldered to metal pads on two circuit boards, which are mounted parallel to one another with an air gap between. Conductive acrylic plates envelope the diodes, providing shielding from the radio-frequency fields produced in the waveguide of the linac. This phantom geometry is more explicitly suited for use with IMRT QA, as one can eliminate any need for angular correction factors to be applied if one keeps the plane of the array face orthogonal to the source of radiation at all times. For IMRT QA, this can be performed by either attaching the array to the collimator during gantry rotation, or by configuring a dummy treatment plan that delivers each field with the gantry set at zero degrees, while leaving the phantom on the couch.

Jursinic *et al.* showed that commercially available surface diodes can have up to ± 12% change in sensitivity as a function of the angle of incident radiation, due to the presence of a copper plane on the circuit board. Diodes used by SNC for their MapCHECK and ArcCHECK phantoms have a similar design, with a copper backing that results in a significant angular dependence. By comparison, prototype diodes with no copper back plane were found to have a ± 3.6% angular dependence [54]. The introduction of an additional layer of copper above the diode's active junction can significantly reduce the angular dependency, but increases perturbation of the radiation field [55].

1.6.2 Phantoms

In the context of radiation therapy, phantoms are structures designed for housing dosimeters at reproducible locations, for physical analysis of treatment plans. They generally consist of materials with composition and density close to that of water, in order to approximate the physical and electron densities of tissues in the patient.

The other phantom used in this thesis for measurement of patient plans was the ArcCHECK phantom (Sun Nuclear Corporation, Melbourne, FL), seen in Figure 1.9. ArcCHECK consists of an

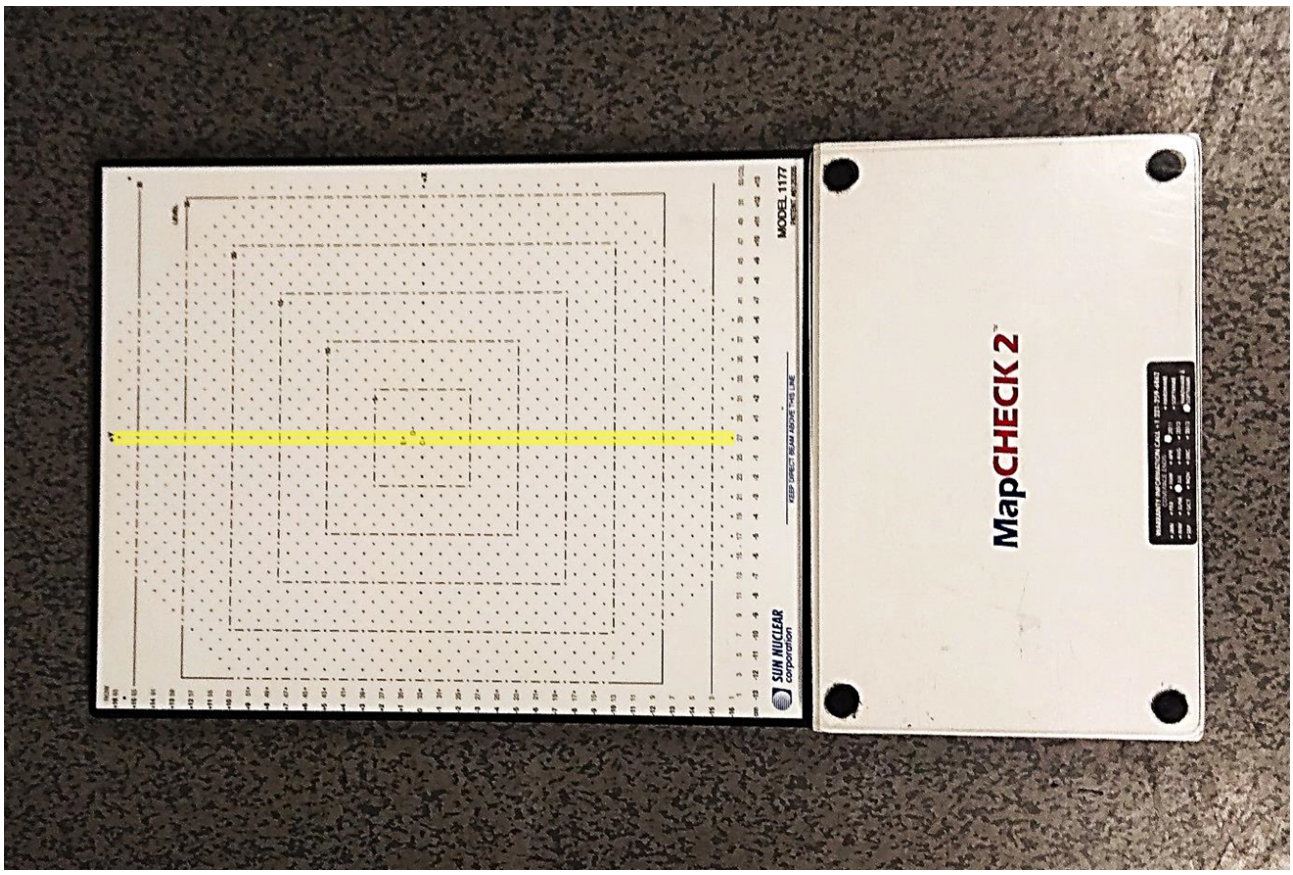


Figure 1.8: Surface of MapCHECK, showing positions of diodes in array. Central diode column is highlighted in yellow. Diodes are housed at a depth of 12 mm below the upper surface of the phantom.

acrylic cylinder, within which are housed 1386 diode detectors in a spiral pattern. By taking the beams for the treatment plan and applying them to a computed tomography (CT) scan of ArcCHECK within the TPS, one can predict the dose you would expect each individual diode to receive were the ArcCHECK placed in the position of the patient during treatment. Comparing the expected dose to the actual delivered dose then provides the physicist with the ability to determine whether the plan is acceptable to deliver.

1.6.3 Gamma analysis

Gamma analysis was introduced as a technique for comparing a reference dose distribution with an evaluated dose distribution. At its most basic level, it measures two functions: the dose difference at each point in the distribution, and the distance-to-agreement (DTA), whereby the evaluated dose distribution is searched to locate the nearest point with the same dose value, and the distance

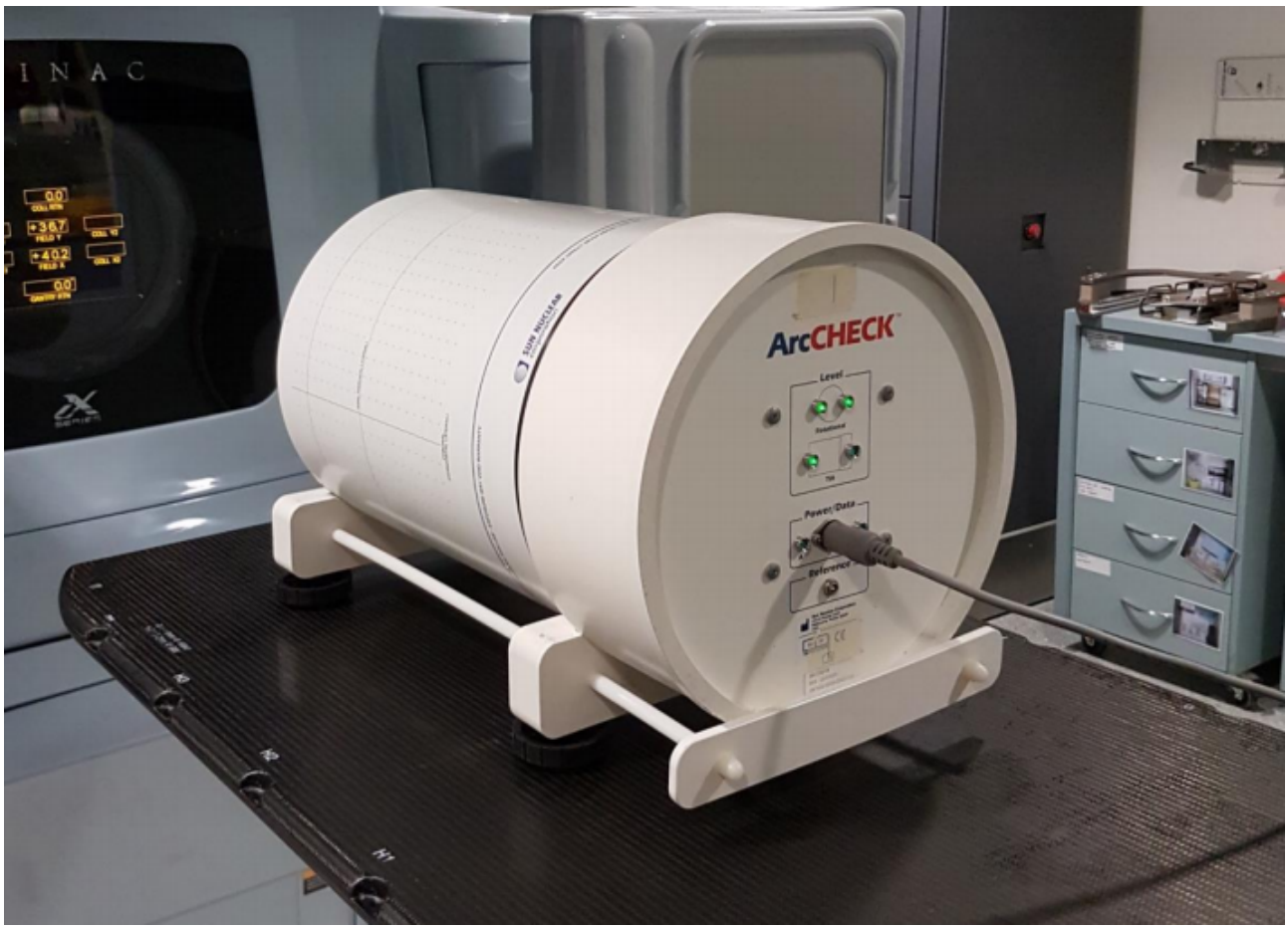


Figure 1.9: ArcCHECK phantom, in position on treatment couch. The hole for ion chamber insertion is at the superior end of the phantom.

between these two points is recorded [56, 57]. The comparison is carried out on a point-by-point basis in a multidimensional spatial frame, in which the calculated and measured dose distributions are overlaid on one another. Acceptance criteria are measured in a space composed of dose and spatial coordinates, with the acceptance criteria forming an ellipsoid surface centred on individual measurement points [57]. If the calculated dose distribution surface intersects with the ellipsoid, then the calculation passes the gamma criteria for that point. A more complete description of the technique is outlined in the seminal paper by Low *et al.* [57].

An example of gamma analysis in a single plane can be seen in Figure 1.10 using the commonly used gamma criteria of 2%/2 mm. It can be seen that the measured dose is slightly colder than the calculated dose at the centre of the distribution. At these points, $\gamma > 1$, indicating that the gamma calculation does not meet the acceptance criteria. For this plan, the gamma pass rate, or number of points meeting the acceptance criteria, was 87.1%.

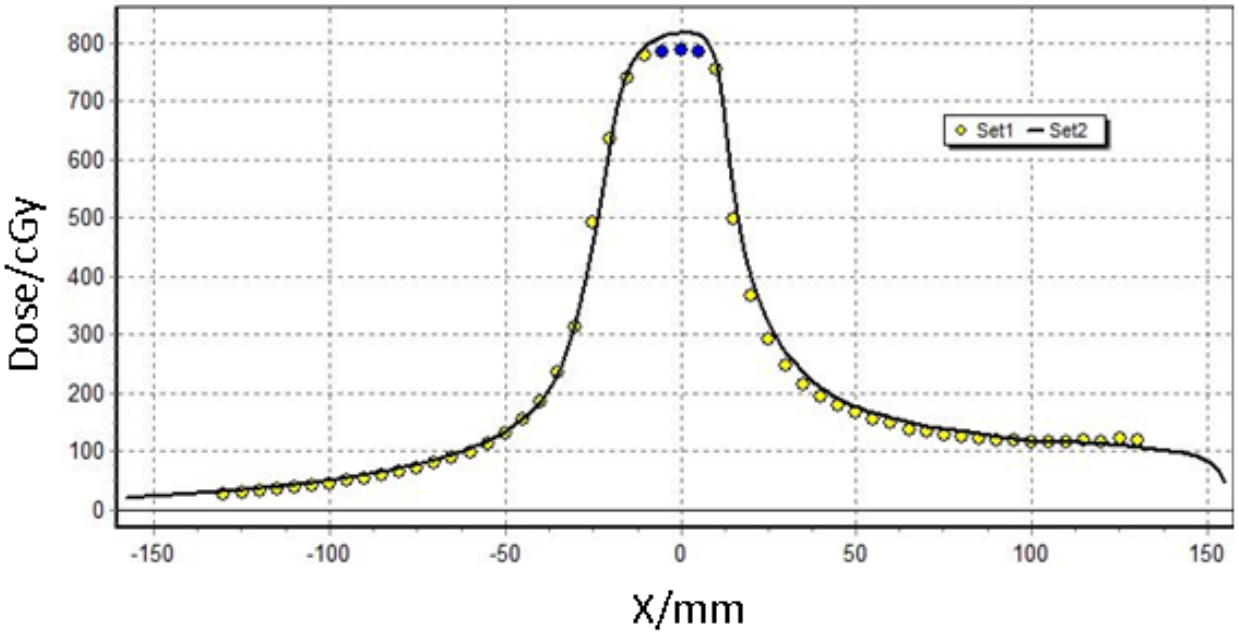


Figure 1.10: Example of gamma analysis in one dimension. Set 1 refers to measured point doses, whereas Set 2 is the calculated reference dose distribution for comparison.

1.7 Receiver Operating Characteristic curves

The Receiver Operator Characteristic (ROC) curve is a tool that can be used to evaluate the sensitivity (SN) and specificity (SP) of discrete classifier tests with binary outcome trees [58], and in particular to compare diagnostic tests and test parameters [59]. This makes the ROC curve a useful tool in radiotherapy for optimization of PSQA processes, which require high SN and SP to be clinically useful. In other words, it is desirable for QA to accurately identify plans that are being delivered incorrectly, while also identifying those plans that are ready to be delivered to patients.

An example of an ROC curve can be seen in Figure 1.11. The curve consists of multiple connected points known as discrete classifiers. The classifier is a diagnostic test which yields two discrete results - for instance, positive or negative. These results can then be classified as TP, TN, FP, and FN (true positives, true negatives, false positives, and false negatives), which can be used to determine the SN and SP of the test. This can be done simply via use of a confusion matrix, as shown in Figure 2.6 in Section 2.5. Hence each discrete classifier generates a (*FP rate, TP rate*) pair which corresponds to a single point in ROC space [60]. By varying the pass threshold of a classifier in a continuous fashion, one can generate an ROC curve similar to that shown in Figure 1.11, since the variation of the threshold will vary each of TP, TN, FP and FN. Fawcett [58] and Lasko *et al.* [60] both provide

excellent introductions to ROC curve analysis, in general and medical settings, respectively.

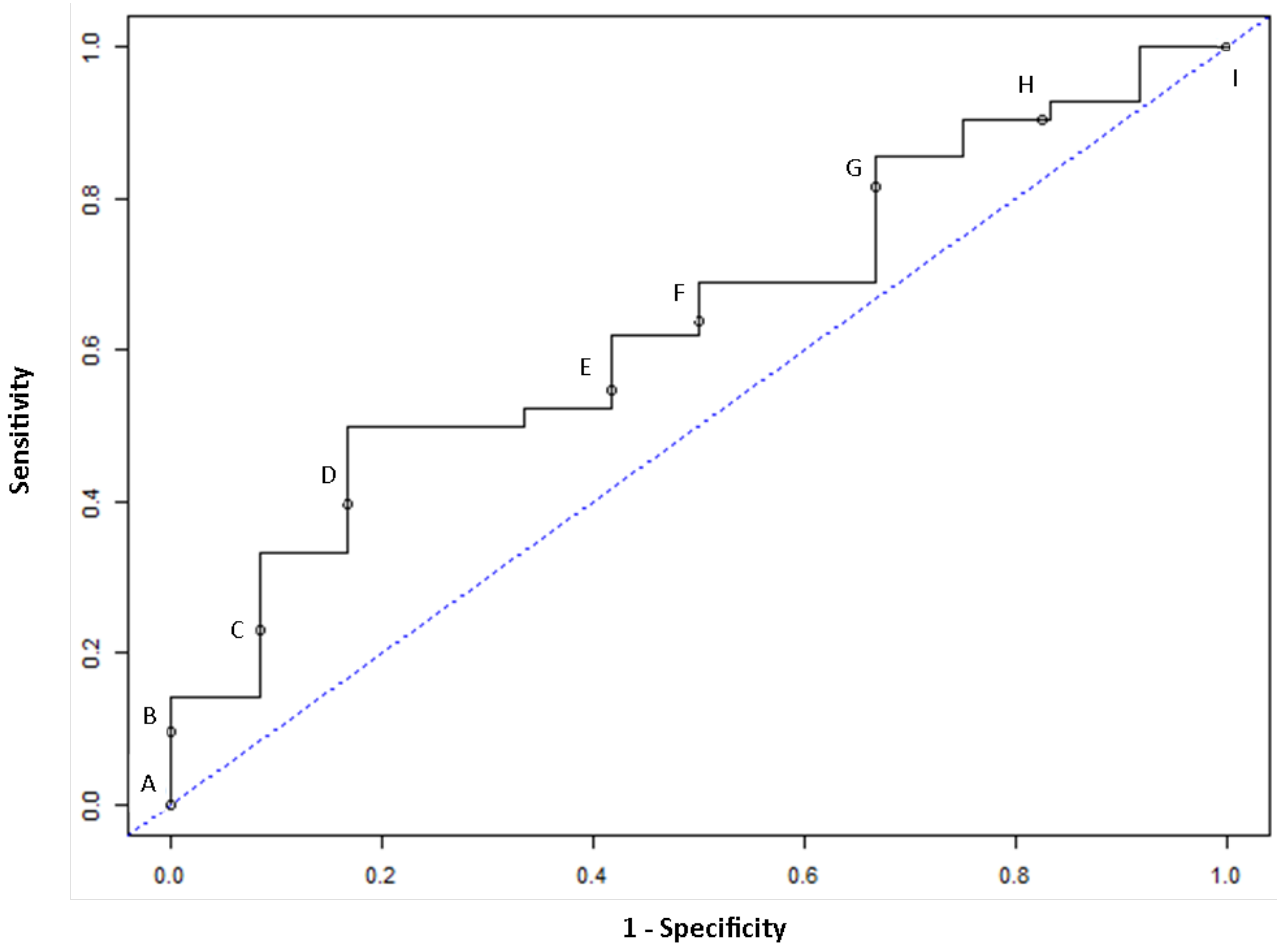


Figure 1.11: Example of an ROC curve. The algorithm for the curve connects the points for classifiers A to I. A test with perfect accuracy would have a curve running vertically from (0,0) to (0,1) and then horizontally to (1,1), while a test that performs no better than random guessing would run diagonally from (0,0) to (1,1), as denoted by the blue dashed line.

Several features in Figure 1.11 are worth noting. Point A at (0,0) is a discrete classifier that represents the strategy of never issuing a positive classification. While this strategy will prevent any false positive errors from being committed, it also gains no true positives. The opposing strategy (always issuing a positive classification) is represented by point I, at (1,1). The blue dashed line represents the strategy of randomly guessing a class. For instance, a classifier at the point (0.7,0.7) would guess the positive class 70% of the time; this strategy would be expected to guess 70% of positive instances correctly but would also have a false positive rate of 70%. The hypothetical point (0,1) would represent the perfect classifier - a strategy that would theoretically produce no false positive errors and classify all positive instances correctly. Hence it can be seen that classifiers (and associated ROC curves) appearing closer to the top-left corner of the graph in ROC space are

preferrable. This would indicate that the classifier is correctly exploiting some information about the class [58]. Any points or curves below the blue dashed line are performing worse than the strategy of randomly guessing a class, which could mean that the classifiers have useful information about the class, but they are applying the information incorrectly [61].

In general, classifiers in the upper right-hand corner of ROC space are considered 'liberal', in that they make mostly positive classifications even with weak evidence, creating a high true positive rate but also generating a high false positive rate. Classifiers in the lower left-hand corner of the ROC graph are considered 'conservative', as they only make positive classifications with strong evidence, meaning they make few false positive errors but their true positive rates are often low [58]. Hence in Figure 1.11, it would be difficult to choose one classifier as being clearly superior to the rest. One could argue that D is closest to the northwest corner of the graph, but since this classifier is quite conservative, it may not be desirable for the particular test being implemented. For instance, if one wishes to label a higher number of cases as being positive (even at the risk of increasing the false positive rate), one may wish to use a more liberal classifier in order to increase their true positive rate. This is a common scenario that is encountered in PSQA, as will be explained later in this thesis.

1.7.1 Indices of accuracy

Visually, ROC curves serve as a useful qualitative analysis of the SN and SP of a diagnostic test. However, there are several quantifiable ROC curve indices that can be calculated in order to compare diagnostic tests.

One of the most commonly used ROC indices, and easiest to interpret, is the area under the ROC curve (denoted as AUC). An ROC curve representing a perfect test has an AUC of 1.0, since the curve will cover the entire chart, while an ROC curve representing a classifier of random chance produces an AUC of 0.5. Hence higher AUC values are desirable, while an AUC under 0.5 either indicates that the test is poorly designed or is interpreting data incorrectly [58]. In the context of PSQA, the AUC has several interpretations - the average sensitivity of the test for all specificity values, the average specificity of the test for all sensitivity values, and the probability that the test will rank a randomly chosen plan that passes relevant tolerances higher than a randomly chosen plan that exceeds said tolerances [60].

In order to calculate an AUC curve from test data, one must derive the true ROC curve from a finite dataset of discrete classifiers. This is analagous to the inference of a continuous statistical distribution from a finite sample [60]. The simplest method of doing this is known as the *empirical* or *nonparametric* method, in which the ROC curve is approximated by connecting each of the discrete points (*SN,1-SP*) linearly. The estimated AUC is then calculated the trapezoidal rule. While this method has the advantage of having no assumptions about the data, it results in an AUC that is biased downward when the number of points on the curve are limited [60].

A weakness of the AUC parameter lies in the fact that a high-AUC classifier can theoretically perform worse in one section of ROC space compared to a low-AUC classifier - i.e. when the classifier curves intersect and cross at one or more points. For instance, if two classifiers must produce a specificity above 80% in order to be useful, then the superior classifier should be chosen based on performance in this useful range. If the classifier curves cross, then the partial area under the curve (pAUC) may be a more accurate measure of classifier performance [60].

With this deficiency in mind, a useful compliment to the AUC is the Youden index, which provides more specific information regarding the efficacy of individual classifiers, rather than the test as a whole. The Youden index can be simply calculated via the following equation:

$$J = sensitivity + specificity - 1$$

where J is the Youden index, also known as Youden's J statistic [62]. The value for the index is between 0 and 1 inclusive, where 0 indicates a test that is no better than random guessing, and 1 represents a perfect test [62]. In the context of PSQA, the Youden index may be used as an aid for determining optimal criteria for the detection of treatment delivery errors.

1.7.2 Thesis objectives

This thesis aims to determine whether MapCHECK in MapPHAN is a viable alternative to film for performing QA on DCAT lung SABR plans. In particular, it aims to assess the efficacy of the phantom at detecting dosimetric errors in OARs, a topic which is scarcely covered in the literature. This will be achieved via ROC curve analysis of introduced errors that were delivered to the phantom, both *in silico* and directly to the MapPHAN on a linac.

Chapter 2 presents the materials and methods used in this study. It details the techniques used for evaluation of errors introduced *in silico* to lung SABR plans, and the methodology for delivering these plans on a linac as a comparison to the *in silico* analysis. It also includes a description of the methods used for commissioning the MapPHAN prior to measurement.

Chapter 3 contains the results generated during this thesis. It links the *in silico* analysis to the phantom measurements via ROC curve analysis, and includes an overview of the results obtained during commissioning of the phantom, with a particular emphasis on the angular dependency corrections performed. The optimal gamma criteria for use in PSQA with this phantom are determined, using the results from the ROC curve analysis.

Chapter 4 is a discussion of the results found in Chapter 3, in which the effect of the angular dependency corrections on the phantom measurements is discussed, and compared with the literature. The statistical accuracy of the ROC curve analysis is scrutinized, along with its potential clinical relevance. Potential areas of future work are also indicated.

Chapter 2

Materials & Methods

This study was comprised of two broad experimental sections. In the first experiment, errors were introduced to lung SABR plans *in silico* on a treatment planning system, in order to calculate the effect of these errors on the dose distributions of each plan. In the second experiment, these plans (with and without errors) were physically delivered to a phantom, and a robustness analysis was performed on the QA results achieved with the phantom. The sensitivity and specificity of MapCHECK in MapPHAN to detect clinically significant errors were determined by analysing the combined results of both experiments.

The broad focus of this thesis was on the sensitivity and specificity of the MapPHAN phantom to introduced errors, with a specific focus on OAR dosimetry, since the QA of PTV doses has already been extensively covered in the literature [63–68].

2.1 *In silico* analysis of introduced errors

Twenty-three patients treated for lung cancer using DCAT lung SABR at Auckland Hospital were randomly selected for inclusion in this experiment. Of these, fourteen patients received 48 Gy to the PTV in 4 fractions, and nine had 60 Gy delivered to the PTV in 8 fractions. All plans were originally delivered using 6 MV flattened X-ray beams on a Varian Clinac iX linac (Varian Medical Systems Inc., CA, U.S.A.). Each plan had been previously optimised in Pinnacle v9.8 or v9.10 (Koninklijke Philips N.V., Amsterdam, Netherlands). Target volumes and OARs were delineated by

the treatment planners and radiation oncologists at Auckland Hospital. Patients were scanned in CT with hands raised above their heads, immobilized in the supine position. The Varian Real-Time Position Management Gating System (Varian Medical Systems Inc., CA, U.S.A.) was used to generate an ITV, via reconstruction of eight breathing phases captured in CT. An isometric margin of 5 mm was then added to the ITV volume to generate the PTV contour. The OARs delineated for these plans included the chest wall, brachial plexus, great vessels, heart, oesophagus, proximal bronchus, ribs, skin, trachea, external, and lungs. Additionally, for some 60 Gy plans, the spinal cord PRV and a volume containing the entire lung minus ITV were contoured. Not all OARs were contoured in each plan, and for certain OARs (such as the spine), only the clinically relevant portion of the OAR was contoured, since any volumes outside of the treatment region would have no significant effect on dose statistics. The planning isocentre was located as close as possible to the PTV centre, within the maximum allowed lateral couch position of 3.5 cm (in order to avoid collisions between the gantry and the couch during gantry rotation). The minimum arc length for each plan was 180^{o}, with typical arc length being 200-220^{o} and each plan being split into either two or three arcs, each with 5^{o} spacing between control points. Beam entry through the contra-lateral lung, spine and heart was avoided where possible.

The plans for these patients were anonymised in Pinnacle, then transferred across to Raystation v5.0 (Raysearch Laboratories, Stockholm, Sweden). This TPS provides efficient tools for the editing of plan parameters and subsequent calculation and comparison of these plans. Prior to transferring these plans to Raystation, the control point spacing for these plans was reduced from 5^{o} to 4^{o} per control point, in order to facilitate the transfer of these plans between the two TPS's. This had the additional benefit of slightly improving the accuracy of the treatment planning calculation, although calculation time increased by 25% as a result of the increase in control point number.

2.1.1 Raystation simulations

Two types of error were introduced to the treatment plans - isocentre shifts and systematic MLC shifts, known as class open errors. Raystation was used to introduce errors in the position of the plan isocentre using incremental errors of 0.2 mm, up to a maximum error of 1 mm, which is the tolerance for disparity between mechanical and radiation isocentre at Auckland Hospital. The isocentre was shifted both ways along all three cardinal axes - inferior-superior, right-left, and

anterior-posterior. This could be considered analogous to shifting the CT dataset in the opposite direction to the isocentre shift. In order to introduce MLC errors, Python 2.7 [69] was used to generate a script that could edit parameters in DICOM (Digital Imaging and Communications in Medicine) files, which are a standard for containing data structures that are commonly utilised in medicine [70]. This script (which can be found, along with the requisite DICOM tags, in Section 1 of the Appendix) was used to introduce errors in the planned MLC positions for each patient, with incremental class open errors of 0.2 mm being added to widen the MLC aperture for all leaves on both MLC banks. For some of the plans, errors were introduced up to a level of 2 mm in order to investigate the expected trends. However, for the majority of plans, the errors were introduced up to a level of 1 mm, as this is the tolerance for MLC position accuracy at Auckland Hospital. As closing of the aperture would have produced a decrease in the calculated dose to all tissues, which is not detrimental to OARs, this type of error was not introduced to the plans.

An example of a class open MLC error is shown in Figure 2.1. It can be seen that any tissue directly beneath the blue sections (close to the PTV boundary) would be most significantly affected by the introduction of these errors, as the removal of shielding allows these sections to be directly irradiated. Tissue adjacent to these sections will also receive an increase in dose due to increased scatter from the blue tissue regions, but the dose contribution from scatter is far smaller than that from direct irradiation.

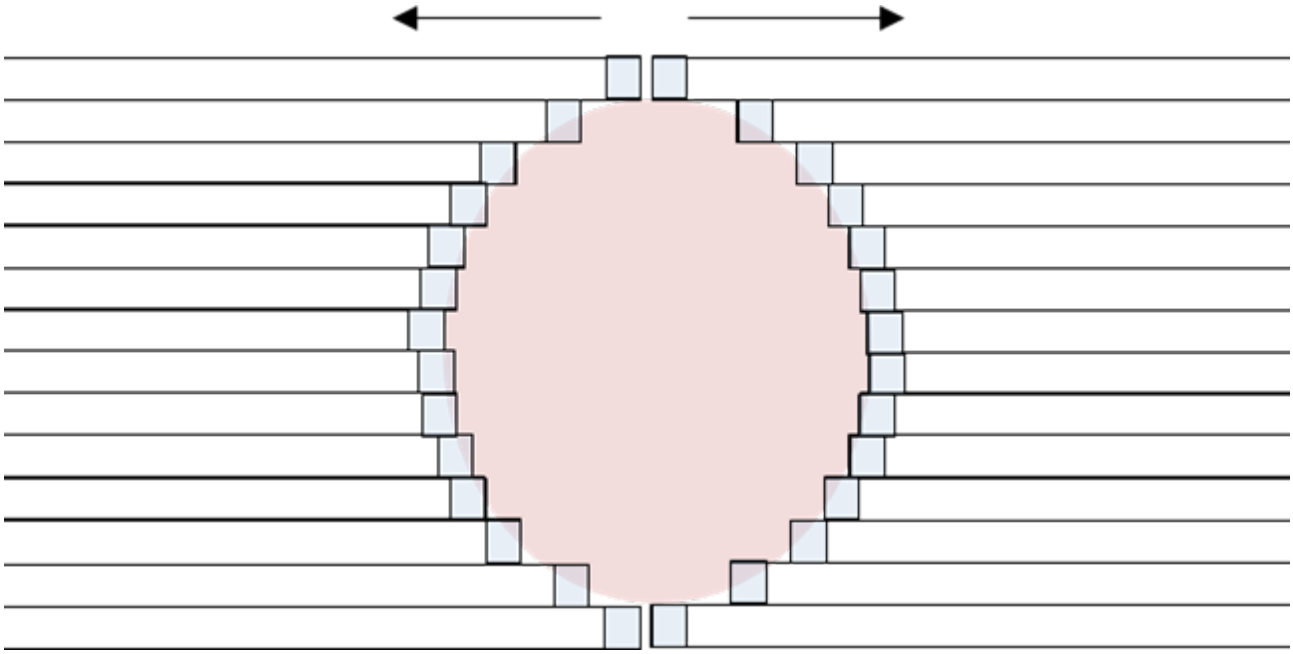


Figure 2.1: Beam's eye view of example PTV (red) beneath MLC leaves for one control point in a conformal arc plan (not to scale). Introduction of MLC class open errors is shown in blue.

Dose distributions for plans, both with and without introduced errors, were calculated in Raystation 5.0, based on the parameters set in the DICOM files. An IGRT couch model was generated in Raystation for each patient, to simulate the position of the couch during actual treatment. A dose grid voxel size of 0.2 x 0.2 x 0.2 cm^3 was used to calculate each plan. DVH metrics were recorded for both the original and edited plans, as shown in Table 2.1. These metrics were chosen based on the constraints used by the planners to optimise the plans, which in turn were based on international recommendations and clinical results, such as the RTOG 0813 lung SABR trial [71, 72].

Table 2.1: Recorded DVH metrics for both edited and original plans.

	PTV dose	
	48 Gy	**60 Gy**
OAR	Planning tolerance	
Brachial plexus		$D_{max} \leq 36$ Gy
Chestwall	$V_{30\ Gy} \leq 70$ cc	$V_{30\ Gy} \leq 70$ cc
External		$D_{max} \leq 78$ Gy
Great vessels	$D_{max} \leq 45$ Gy	$D_{max} \leq 54$ Gy
Heart	$D_{max} \leq 30$ Gy	$D_{max} \leq 44$ Gy
Oesophagus	$D_{max} \leq 27$ Gy	$D_{max} \leq 40$ Gy
Proximal bronchus	$D_{max} \leq 30$ Gy	$D_{max} \leq 44$ Gy
Spinal cord (PRV)		$D_{max} \leq 28$ Gy
Trachea	$D_{max} \leq 30$ Gy	$D_{max} \leq 44$ Gy
Whole lung - ITV		$V_{20\ Gy} \leq 15\%$
Whole lung	$V_{5\ Gy} \leq 60\%$	
Ribs	$D_{max} \leq 40$ Gy	
Skin	$D_{max} \leq 36$ Gy	

The maximum point dose for each contoured organ within the plan was calculated in Raystation and recorded. These values were plotted as a function of both the incremental MLC and isocentre errors. The gradients of these plots were then used to determine the significance of the introduced errors.

2.2 Commissioning of MapPHAN

The phantom used in this thesis was the MapCHECK 2, Model 1177. This was mounted between two slabs of solid water which comprised part of an old MapPHAN accessory, previously designed for use with the MapCHECK 1 phantom (Model 1175). To fit the MapPHAN accessory to the MapCHECK phantom, two solid water inserts with dimensions of 11 mm x 29 mm x 300 mm were

fitted between the two MapPHAN slabs. This was done so the inserts abutted the left and right sides of the MapCHECK phantom, as seen in Figure 2.2.

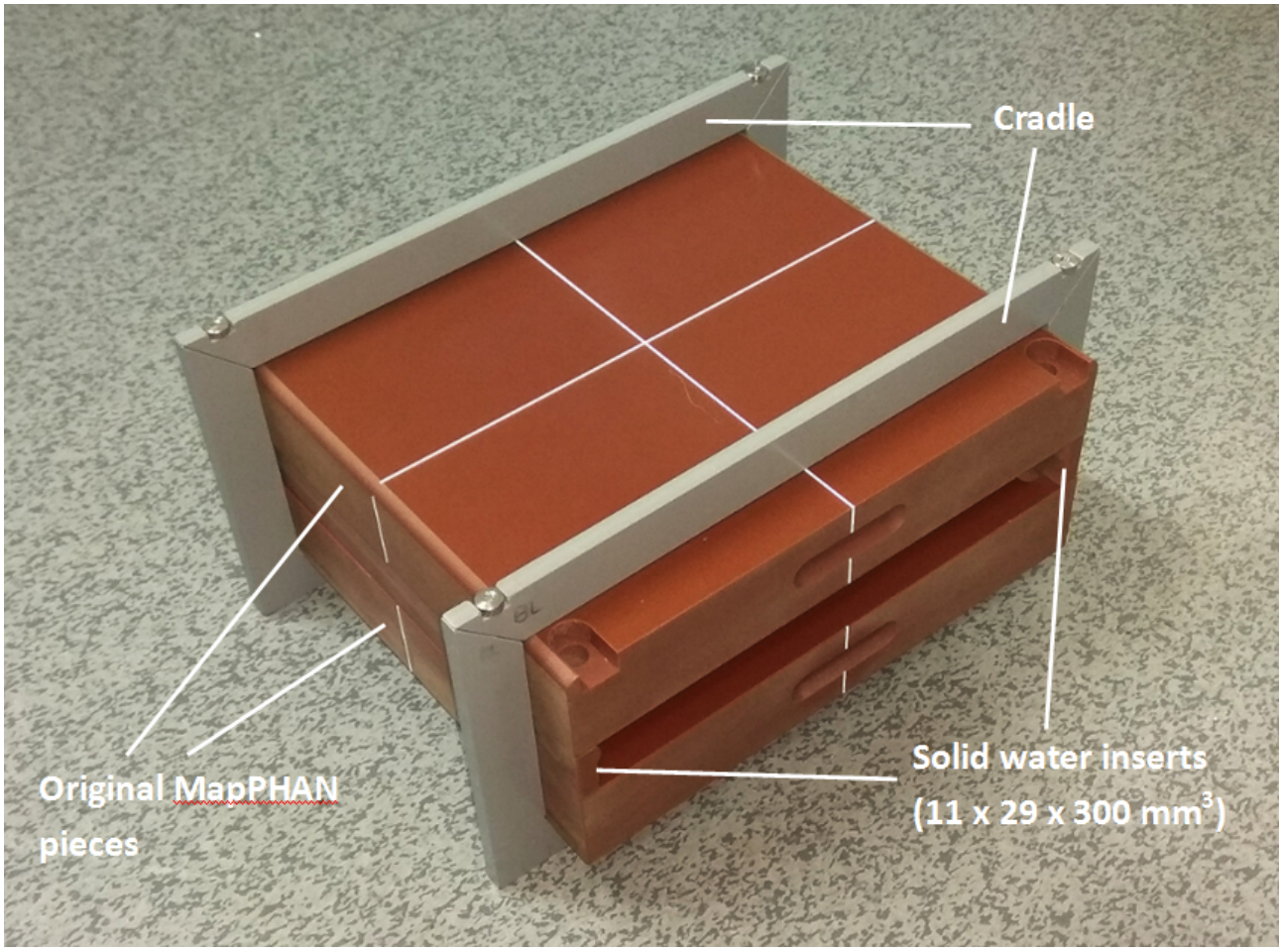


Figure 2.2: Image of the MapPHAN setup used in this thesis, without MapCHECK inserted. Only the top and bottom layers of solid water were from original MapPHAN phantom. The entire phantom can be disassembled by removal of the screws in the cradle.

Sun Nuclear provides the MapPHAN as a solution for the QA of rotational arc therapy treatments. However, there are caveats to this statement. The documentation accompanying the accessory [73] states that the phantom should be placed in a coronal orientation when the majority of dose being delivered is incident from either anterior or posterior beams at G0 and G180. Likewise, if the majority of incident radiation is coming from a lateral direction (around G90/G270), then the phantom should be oriented in the sagittal position. The manual states that this practice 'minimizes angular dependence and maximises the dose map measurement area'. The manual goes on to state that 'Compared results in Relative Dose mode may display a low pass rate' if the dose is normalised to the point of maximum dose, and this point varies significantly from other points. The probability of this occurring 'increases when the MapCHECK/MapPHAN assembly is irradiated from an angle

other than zero degrees'. These statements throw into question the validity of using this device for any sort of rotational delivery, unless additional steps are taken to minimise the angular dependence of the device. This has been performed before using various methods [74–76], but due to the unique geometry of the in-house MapPHAN setup used in this experiment, it was decided to analyse the angular dependencies of the phantom in full detail.

The use of 5 g cm^{-2} of buildup material is recommended by Sun Nuclear for two main reasons: to achieve higher precision in high dose gradient regions, and to minimize scatter radiation to the instrument electronics [73]. The phantom used in this thesis only provides 2 g cm^{-2} of water-equivalent buildup along its lateral sides, with 5 g cm^{-2} provided on the top and bottom surfaces. However, this was deemed appropriate for use in 6 MV beams, as it provides greater depth than the 1.5 g cm^{-2} required to avoid measuring in the buildup region of the 6 MV PDD (not including the additional plastic on the sides of MapCHECK itself). While scatter radiation is, in theory, an increased factor in the measurement of lateral beams for this phantom, the proportion of control points with beams passing directly through this section of the phantom is relatively low for full conformal arc treatments. Additionally, the shorter range of scattered radiation makes it more likely to be absorbed by diodes on the periphery of MapCHECK's array. During the course of this thesis, it was found that the dosimetrically valid portion of the phantom was comprised of the diodes within 4 cm laterally of the central phantom diode column, based on the angular dependency measurements taken in Section 3.2. Hence it was assumed that scatter radiation had a minimal effect on measurement, as the majority of scattered radiation would be absorbed prior to reaching the diodes at this depth in the phantom. This is discussed further in Section 3.2.

The MapCHECK phantom itself had previously been commissioned for use within the department, using documents such as the AAPM TG40 report [77] as guidelines for IMRT commissioning standards, among others. This meant that parameters such as dose rate dependency, scatter response, dose reproducibility and dose linearity had already been investigated and controlled for. However, the phantom had only been previously commissioned for receiving beams orthogonal to the array plane, where the phantom was always placed in the same orientation relative to the gantry. This limits the device to performing what is essentially a fluence-based measurement for IMRT plans. MapPHAN, on the other hand, is designed to allow the MapCHECK phantom to be irradiated from all gantry angles. This allows measurement of plans delivered using arc therapy techniques.

In order to complete the commissioning process, four key operations needed to be performed:

- Quantify the angular dependency of the diodes within the phantom;
- Model the phantom (MapCHECK housed within MapPHAN) with appropriate physical and electron density overrides in Raystation, in order to correctly simulate the dose delivered to the phantom for each plan;
- Perform an absolute dose calibration on the central diodes of the phantom;
- Perform end-to-end testing of the phantom with both generic conformal arc beams and patient plans, in order to test the accuracy of the dose calibration and density overrides.

The methodology of how the above were performed is outlined in the following subsections.

2.2.1 Angular dependency of diodes and determination of density overrides

The phantom was scanned using a Philips Somatom Sensation CT scanner, in both coronal and sagittal orientations. The scans were then imported into Raystation, where streak artefacts were noticed in the plane of the detectors, as can be seen in Figure 2.3. In addition to the streaks in the detector plane, a secondary streak pattern was observed radiating outwards from the diode plane, whereby the CT reconstruction algorithm calculated the density of materials closer to the array as being denser than those at a greater distance. This was confirmed by obtaining readings of the Hounsfield units measured at different points in the solid water, as viewed in Raystation. These artefacts have been reported previously by multiple authors [74–76]. Keeling *et al.* performed a similar analysis of the percent dose difference between measured and planned doses with no density overrides applied, reporting up to a 38% discrepancy between measured and calculated doses for 6 MV X-ray beams with 5 x 5 cm^2 fields [74]. The dependency was slightly lower for a 10 x 10 cm^2 field, and significantly lower for similar measurements performed with 10 MV, presumably due in part to the fact that the attenuation coefficients of the diode materials decrease as the photon energy increases.

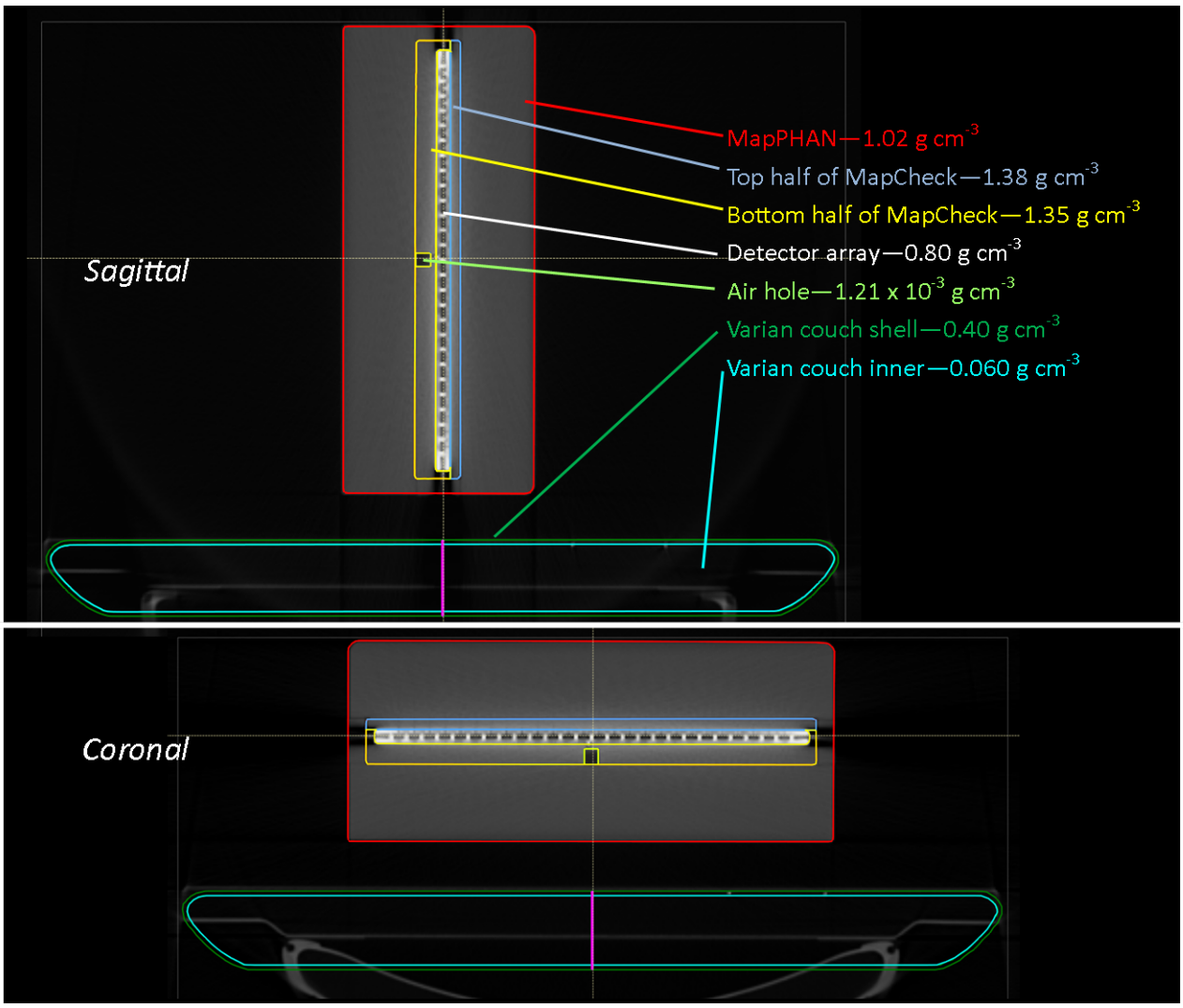


Figure 2.3: Transverse slices taken from CT scans of MapPHAN in sagittal and coronal orientations, with mass density overrides applied. A Varian couch model has been included. Note the streak artefacts in the plane of the diode array. All material contours were generated with an electron density equal to that of water, as per Sun Nuclear recommendations for the standard MapPHAN.

The angular dependency is worst at gantry angles of 90° and 270°, where the beam axis is parallel to the detector array. This increases the attenuation of the beam and accentuates the inherent angular dependencies of the diodes themselves [74]. The increased attenuation of the beam in the plane of the diodes can be observed in Figure 2.3, where dark streak artefacts are seen in the plane of the diode array, which contains circuit boards with high atomic number materials that significantly attenuate low-energy X-rays produced by CT scanners. This effect has also been observed by Jursinic *et al.*, who used a MV CT scanner with higher energy X-rays that aren't preferentially attenuated to the extent that kV X-rays are. Use of the MV scanner reduced the streak artefacts to the point that the density of the phantom could be derived directly from the CT scan [75]. As no

MV CT scanner was available at Auckland Hospital, another method was necessary to determine the densities of the materials housed within the phantom.

Lacking the ability to perform MV CT imaging, it was decided that the best alternative would be to apply density overrides to all sections of the phantom (including materials of known density) in order to reduce the dependency to acceptable tolerances. Using the scanned phantom model in the coronal orientation, density overrides for the phantom's contours were determined by using the angular dependency measurements with the phantom in the coronal orientation. These overrides were then applied without modification to the phantom model in the sagittal orientation.

Jursinic *et al.* [75] claim to have reduced the angular dependency of the MapCHECK phantom from ± 20% to within ± 2%. Based on their MV CT images, they were able to deduce that the mean density of their MapCHECK phantom was 1.21 g cm^{-3}, while the mean density of the solid water in MapPHAN was 1.05 g cm^{-3}. Using these applied densities, they measured the resultant angular dependency of the phantom using 25 x 25 cm^2, 6 MV X-ray beams delivered to the central diode of MapCHECK, which was aligned with linac isocentre. These beams were delivered at 15 degree increments through all gantry angles. Both measured and calculated doses were recorded for the central diode, and the ratio taken to determine the accuracy of the calculated doses. Due to the lack of a suitable MV CT scanner for accurate measurement of the array densities, it was decided that optimising the phantom densities in an inverse manner would be the most practical solution. By performing angular dependency measurements and comparing the dosimetric results to those calculated in Raystation, the appropriate mass and electron densities of the phantom materials could be optimized in Raystation.

The overrides were optimised in an iterative fashion. With the phantom placed at isocentre, 10 x 10 cm^2 fields of 100 MU, 6 MV X-ray beams were delivered to the phantom at static gantry angles through the full range of gantry angles (shown in Figure 2.4) at increments of 10° for each beam. (The choice of 10 x 10 cm^2 represents a standard field size that provides close to full scatter conditions for the central diode column, while also being close to the relative size of the fields used for SABR.) Raystation was used to calculate the expected dose to the five central diodes of the phantom for each delivered beam. The ratio of the measured and expected doses for the mean dose recorded by the 5 most central diodes was then calculated for each beam, after normalisation to the G0 measurement. From these ratios, a chart of the angular dependence based on the relative dose

obtained at each gantry angle was generated. As an additional analysis tool, the dose distributions for both the measured and computed beams were compared using SNC Patient software (Sun Nuclear Corp., Melbourne, Florida). By changing the material overrides on the simulated phantom in an iterative fashion, the agreement between measured and calculated doses at the centre of the diode array was optimized. By taking the mean dose from the five central diodes, rather than the dose from the central diode only, the statistical uncertainty in these measurements was reduced. Since the maximum lateral separation between the central five diodes is 1 cm, it can be assumed that the dose falloff over this region from lateral beams is approximately linear. Hence by taking an average of the doses measured in this 1 x 1 cm^2 region, one can obtain an accurate assessment of the dose at the central diode, with better statistics than would have been obtained from a single point dose measurement.

2.2.2 Conformal arc measurements

Following the application of density overrides, conformal arc beams were delivered to MapPHAN to determine the accuracy of the overrides. Three 360° arc beams of 600 MU each were delivered to the centre of the coronal phantom, with 10 x 2 cm^2, 10 x 5 cm^2, and 10 x 10 cm^2 field sizes respectively. The same arcs were simulated in Raystation. The doses received by the MapCHECK array were imported into SNC patient software, which produced a visual readout of the point doses on the array. These were then compared to their respective Raystation-calculated dose distributions.

Additionally, film measurements were taken to determine the actual field size as set by the jaws during delivery of these arcs, in order to ensure that the output of the linac matched that being simulated in Raystation. Film was placed at the level of isocentre, with 5 cm of solid water buildup placed on top of the film, and another block of solid water placed underneath the film. 6 MV X-ray beams were delivered at G0. The field sizes were then determined as the average distance between the 50% isodose lines in the beam profiles produced by SNC, in accordance with how the jaws were originally calibrated. All field sizes were within ± 2% of the expected results.

All conformal arc plans had a collimator offset of either 10° or 350°, which eliminated the MLC tongue-and-groove effect. This occurs when diode detectors aligned in the plane of the MLC's inter-leaf gaps receive extra dose as the gantry rotates in the plane of these gaps, due to the reduced

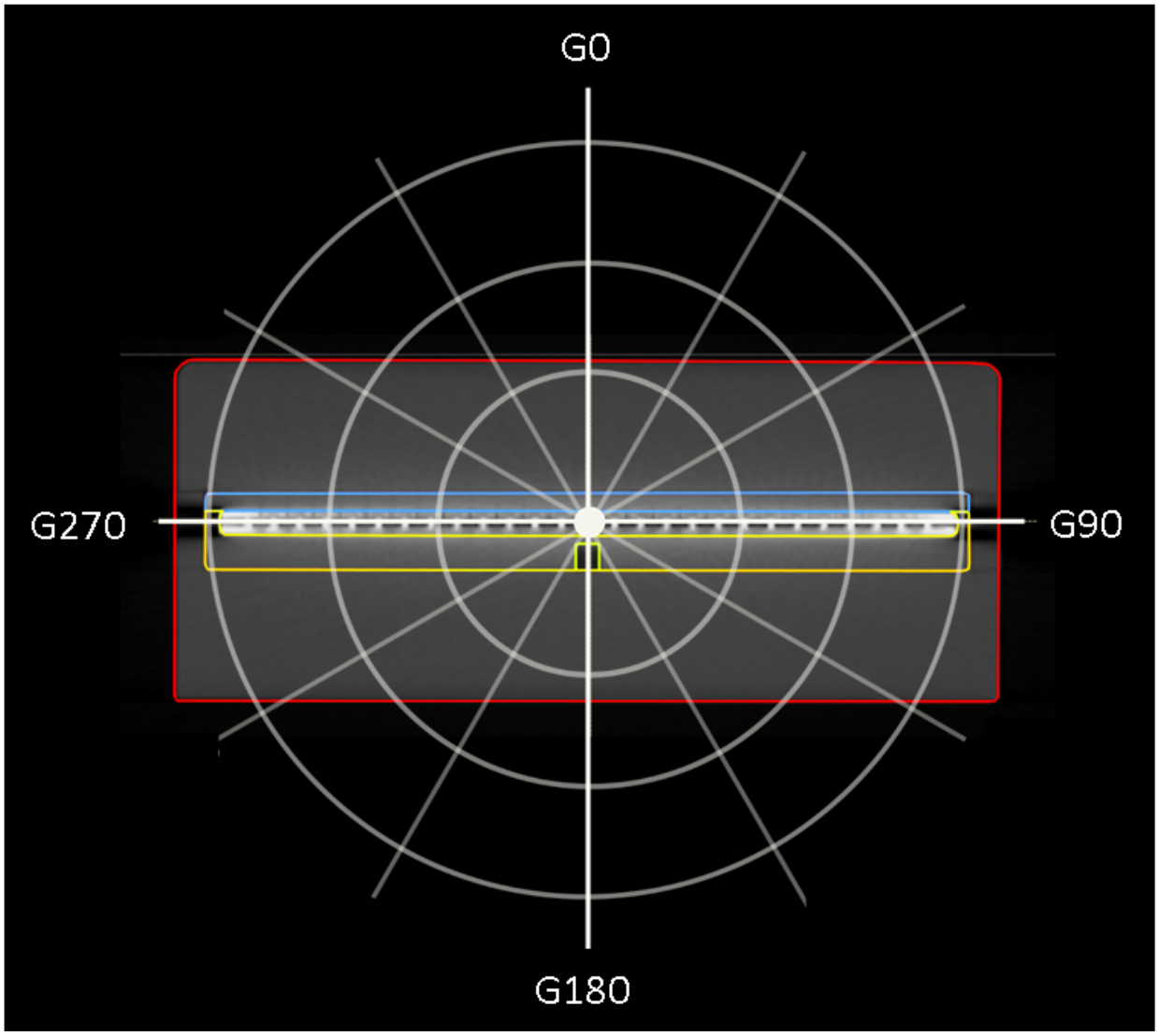


Figure 2.4: Raystation contours of MapPHAN in coronal orientation, with gantry angle coordinates overlaid.

shielding provided by these gaps. This effect has been observed to decrease the gamma pass rate for plans without collimator rotation that were checked with MapCHECK 2, if not accounted for [78].

2.3 Patient specific quality assurance measurements

This section outlines the methodology behind all physical measurements performed on the linac. This includes the measurement of both the original and edited patient plans on MapCHECK in MapPHAN, along with the validation measurements performed for these plans, performed with film and ionization chamber measurements in the ArcCHECK phantom.

2.3.1 MapPHAN measurements

Once commissioning of the MapPHAN was completed, the template for the phantom with appropriate density overrides was saved in Raystation for use in PSQA. The plan parameters for the original patient plans were applied to this template to simulate the expected dose distributions on this phantom, which were then saved as DICOM RTPlan and RTDose files. The treatment plans (both original and with MLC and isocentre errors applied) were then physically delivered to the phantom using the same Varian Clinac iX linac. From the 23 patients selected for this study, 10 were selected for linac measurements. The original DICOM plan files, along with the plans with MLC edits included, were all transferred electronically to the linac in order to be delivered physically. Isocentre shift plans were delivered by physically shifting the patient couch to simulate misalignments between the machine isocentre and planning isocentre, while MLC errors were introduced by the modified DICOM files for each patient plan, which were exported directly from Raystation. Of these 10 patients, one was excluded from measurement due to the necessity of introducing couch shifts to correctly align the dose distribution with the phantom isocentre. This prevented the introduction of an unnecessary confounding factor. The parameters for these plans were used to calculate dose distributions on the simulated MapPHAN phantom in Raystation, using both the original plans, and plans with MLC and isocentre shifts included. The RTPlan and RTDose DICOM files (containing the plan parameters and plan dose data, respectively) for these virtual plans were exported from Raystation and imported into SNC Patient. This software allows the virtual plans to be compared to the actual dose delivered to the phantom using gamma analysis.

The physical phantom was aligned with the mechanical isocentre of the linac using the in-room lasers, as shown in Figure 2.5. Couch shifts were simulated on the MapCHECK phantom model in Raystation for each plan, in order to align the region of maximum dose with the central diode of the MapCHECK phantom. These couch shifts were then replicated when performing physical measurements on the linac. This was done for two reasons. Firstly, the region close to the point of maximum dose is in a low dose gradient region. This allows the physicist to take point dose measurements with greater accuracy at the centre of the dose distribution; any sub-millimetre misalignments will shift the point of measurement within this homogeneous dose region, with negligible effect on the measurement. Secondly, the angular dependency measurements revealed that measurements taken within 5 cm of the central diode were most accurate; this is outlined in detail in

Section 3.2.

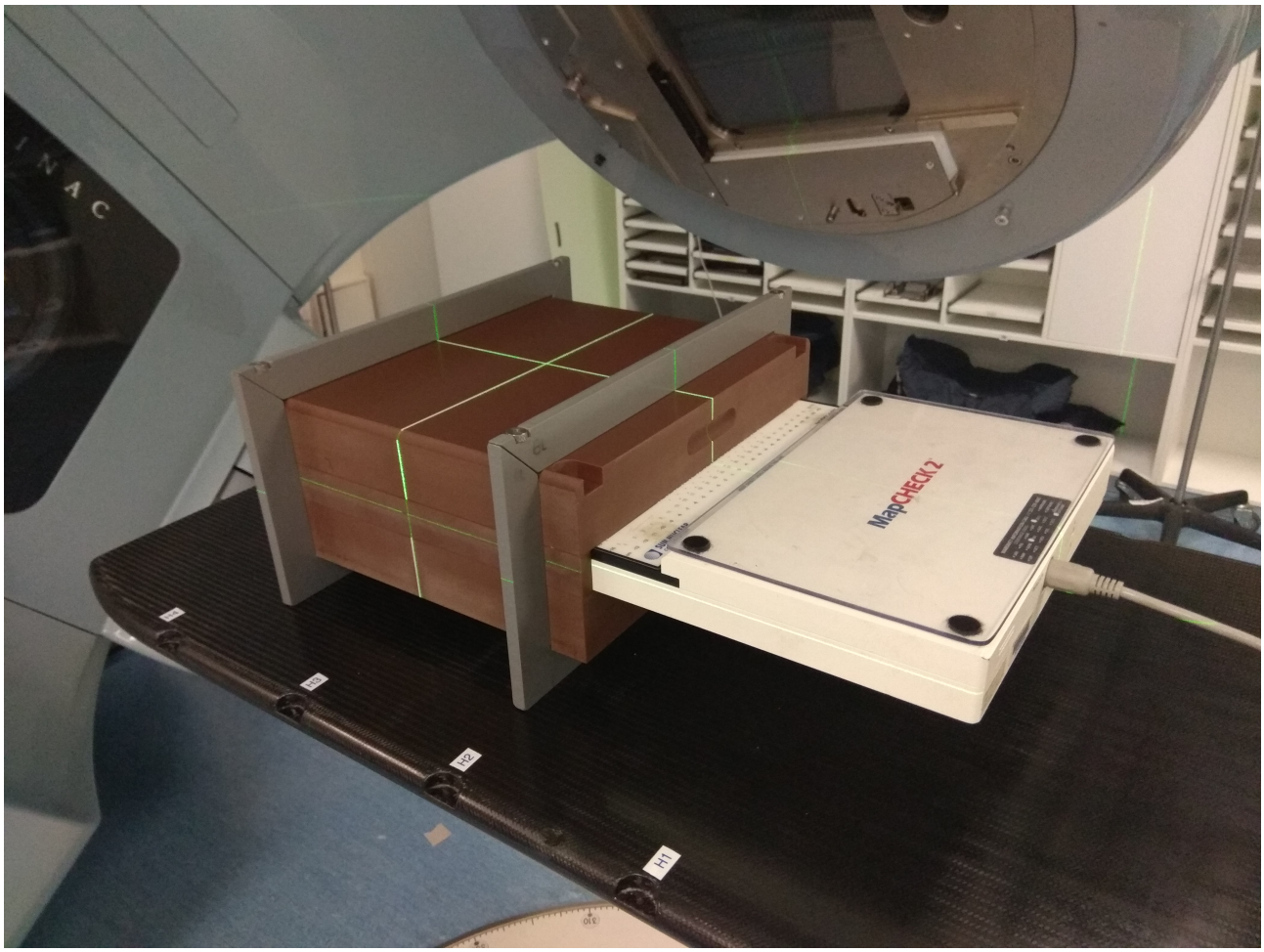


Figure 2.5: Setup used for MapCHECK inserted into MapPHAN, placed on the linac couch. The linac head can be seen at the top right hand corner. The central diode in the phantom was aligned with the machine isocentre by aligning the isocentric lasers with lines painted on the phantom.

MapCHECK has a dose calibration tool which allows the phantom to correct for daily fluctuations in linac output. The phantom is set up under conditions with a known central axis dose, which is calculated in Raystation prior to delivery. This can theoretically be performed for any level of buildup, as long as it is compensated for in Raystation. Prior to all measurements, 100 MU of 6 MV X-rays were delivered to the phantom with the central diode positioned at isocentre, in order to calibrate the device. Initially, the device was calibrated using a single solid water slab which provided 3 cm of water-equivalent buildup material. Following analysis of the plans, the device was calibrated within the MapPHAN accessory (which also consists of solid water), which provided 5 cm of water-equivalent buildup. The choice of calibration technique had a significant effect on the outcome of the experiment, and this is discussed in the Results section.

2.3.2 ArcCHECK measurements

In order to validate the quality of the plans being delivered to MapCHECK, a near-identical process was performed to generate simulated dose distributions on the ArcCHECK phantom created by SunNuclear. ArcCHECK has a film insert which can be used purely for film measurements without measuring the dose in the diodes. It also has an insert with a hole for insertion of a CC04 ionization chamber at the centre of the phantom. This setup can then be used to perform point dose measurements for each QA plan. For this project, the diode array in ArcCHECK was not used for measurements, with only point dose measurements being taken.

Point dose measurements

Ion chamber measurements were performed for the nine original plans that were delivered to MapCHECK, using a 0.04 cc Scanditronix/Wellhofer CC04 ionization chamber (IBA Dosimetry, Schwarzenbruck, Germany) coupled with a UNIDOS E Electrometer (PTW, Freiburg, Germany). For these measurements, a PVC insert was placed inside the ArcCHECK phantom, and the chamber placed inside the centre of the insert. As with the MapPHAN measurements, the plan was calculated on the CT dataset containing the simulated version of the phantom in Raystation, with couch shifts applied to ensure that the maximum dose region of the dose distribution was centred on the ion chamber position.

Film measurements

As a further validation of the plans being delivered to MapCHECK, film QA was performed on three of the nine original plans, using a film insert which was placed in the centre of the ArcCHECK phantom. As with the ion chamber measurements, the simulated dose distribution was aligned with the central part of the film, using couch shifts which were then replicated on the linac.

2.4 Analysis of MapPHAN measurements

Analysis of the measured plans involved comparing the measured dose distribution in each respective phantom to the simulated dose distribution generated by Raystation, using SNC Patient software. This takes point doses in the measured dose distribution (in the case of MapCHECK, doses from each diode; in the case of film, a grid is applied with dose measured at regular spatial intervals) and compares them to the continuous dose distribution surface plot generated by Raystation. Gamma analysis was then performed to obtain a quantitative analysis of the affinity between the measured and simulated dose distributions.

Six gamma analysis settings were used: 1%/1 mm, 2%/1 mm, 2%/2 mm, 3%/2 mm, 3%/3 mm, and 5%/1 mm. These gamma criteria are commonly used clinically, and are pertinent to SABR due to their low distance specifications, all being less than or equal to 2 mm, which is representative of the stringent setup margins required for accurate SABR delivery. As per ADHB protocol, shifts were automatically applied in SNC Patient to align the measured and calculated dose distributions in an optimal fashion. This is done to account for setup error on the part of the physicist. Following application of shifts, the dose recorded in the central diode of the phantom was tabulated for each plan, as well as the disparity between the recorded and calculated central diode dose.

For all measurements, a gamma threshold of 10% was set, meaning that any points receiving dose below 10% of the maximum point dose were not included in the final gamma calculation. This is normal clinical practice, in order to exclude points in low dose regions where the dosimetric uncertainty is highest. Since none of the DVH metrics for OARs (as shown in Table 2.1) were affected by volumes receiving doses below 10% of the PTV dose - 48 Gy or 60 Gy - this is acceptable practice.

A 0.8% measurement uncertainty was applied to all absolute dose comparisons, which consisted of the following errors defined by SNC:

- Measurement electronics error: 0.08%
- Array calibration uncertainty: 0.5%
- SSD uncertainty (2 mm at 100 cm SSD, combining QA and calibration setup errors): 0.4%

- LINAC output fluctuation: 0.15%

- Diode temperature dependence (based on room fluctuation and array gradients of 1°C: 0.5%

These errors were then combined in quadrature according to the GUM (Guide to the Expression of Uncertainty in Measurement) [79] method of analysing errors to obtain a standard uncertainty of 0.8%, with coverage factor of 1 [80]. This was added to the percent tolerance in the gamma analysis. For instance, the 1%/1 mm criteria is actually calculated by using cutoff criteria of 1.8% dosimetric difference and 1 mm shifts. No angular dependency correction was included in these errors, despite the fact that SNC designed MapCHECK for use in the MapPHAN phantom, and includes this uncertainty (0.5%) in measurements made for ArcCHECK.

2.5 ROC curve analysis

In order to determine the sensitivity and specificity of the MapCHECK phantom to the introduced plan errors, ROC curves were generated using the 'ROCR' package created by Sing *et al.* [81] in the statistical programme R [82]. The classifiers used in this thesis were based on the confusion matrix shown in Figure 2.6. Several common metrics can be calculated from this matrix, including:

- the true positive rate (AKA sensitivity or recall) of the classifier, estimated as

$$sensitivity = \frac{TP}{P}$$

- the false positive rate of the classifier, estimated as

$$fp\,rate = \frac{FP}{N}$$

- the specificity of the classifier, estimated as

$$specificity = \frac{TN}{FP + TN} = 1 - fp\,rate$$

- the accuracy of the classifier, estimated as

$$accuracy = \frac{TP + TN}{P + N}$$

- the precision of the classifier, estimated as

$$precision = \frac{TP}{TP + FP}$$

From these metrics, the ROC curve itself is built, with TPR (or sensitivity) on the ordinate and FPR (or 1 - specificity) on the abscissa.

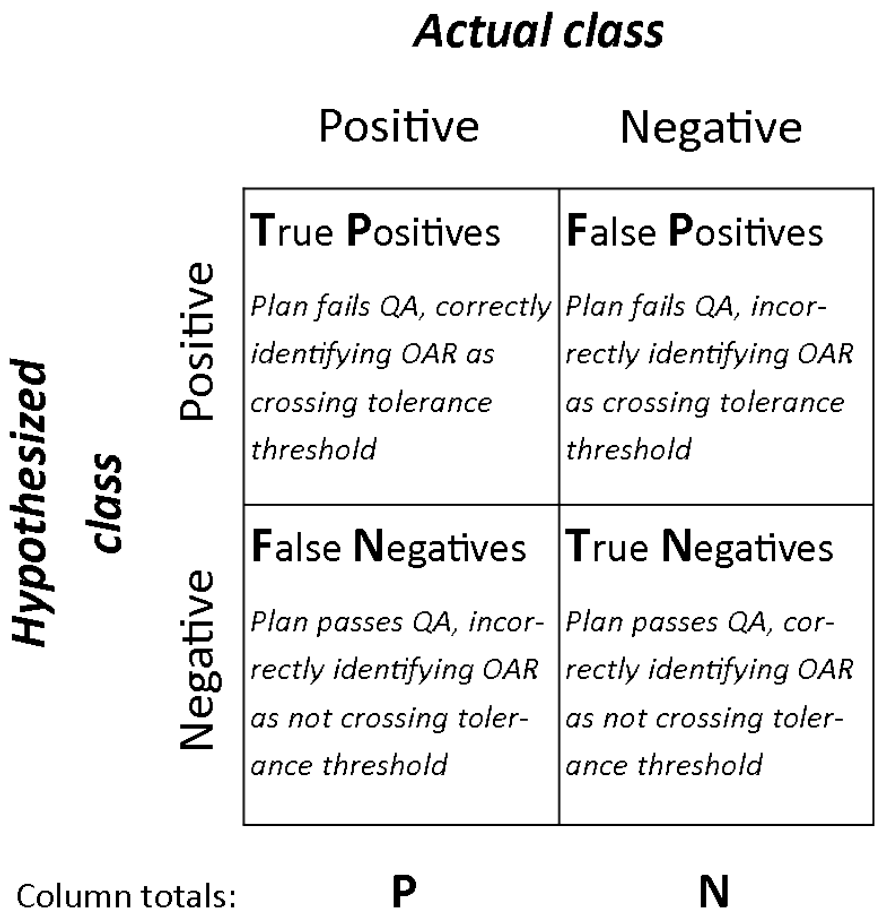


Figure 2.6: Confusion matrix (also known as a contingency table) used for generation of ROC curves specific to OARs in this thesis.

In the context of Figure 2.6, only plans that had OAR tolerances exceeded in Raystation as a direct result of introducing an error had a hypothesised class of positive. That is, if a fully optimised plan with no introduced errors already included an OAR that was receiving in excess of a tolerance dose,

then this plan should still have passed QA (with hypothesised class being negative), assuming it was being delivered as intended. The assumption was made that any Type A errors generated during the delivery of certain plans were negligible compared to the errors being deliberately introduced.
Following generation of the curves, the AUC and Youden index were used to determine the optimum gamma criteria to be used for PSQA.

Chapter 3

Results

3.1 *In silico* analysis of introduced errors

3.1.1 Dosimetric comparison between edited and original plans

As a preliminary analysis, a dosimetric comparison was performed between the original plans and the plans with the most substantial errors introduced, using the dose difference tool in Raystation's plan evaluation module. This was performed as a qualitative analysis to reveal the regions with the greatest dosimetric fluctuations relative to the position of the PTV, upon introduction of errors to the plans. The comparisons were performed for both the conformal arc plans and VMAT plans.

Figures 3.1 and 3.2 show the dosimetric effect produced by 1 mm MLC opening and 1 mm isocentre shifts for one patient plan, which was indicative of the MLC results in general. The results, when combined with the analysis data in Figures 3.3 and 3.4, provide some insight into the general regions that are most likely to be influenced by introduced errors. For MLC deviations, the effect is most pronounced in the region directly beneath the MLC positional shift, as traced from the virtual source of the beam to the point in the patient. This is as expected, since a variation in MLC position is effectively an addition (leaf closing) or subtraction (leaf opening) of beam shielding along the ray line.

As can be seen in Figure 3.2, shifting the isocentre anterior resulted in a dosimetric shift that was

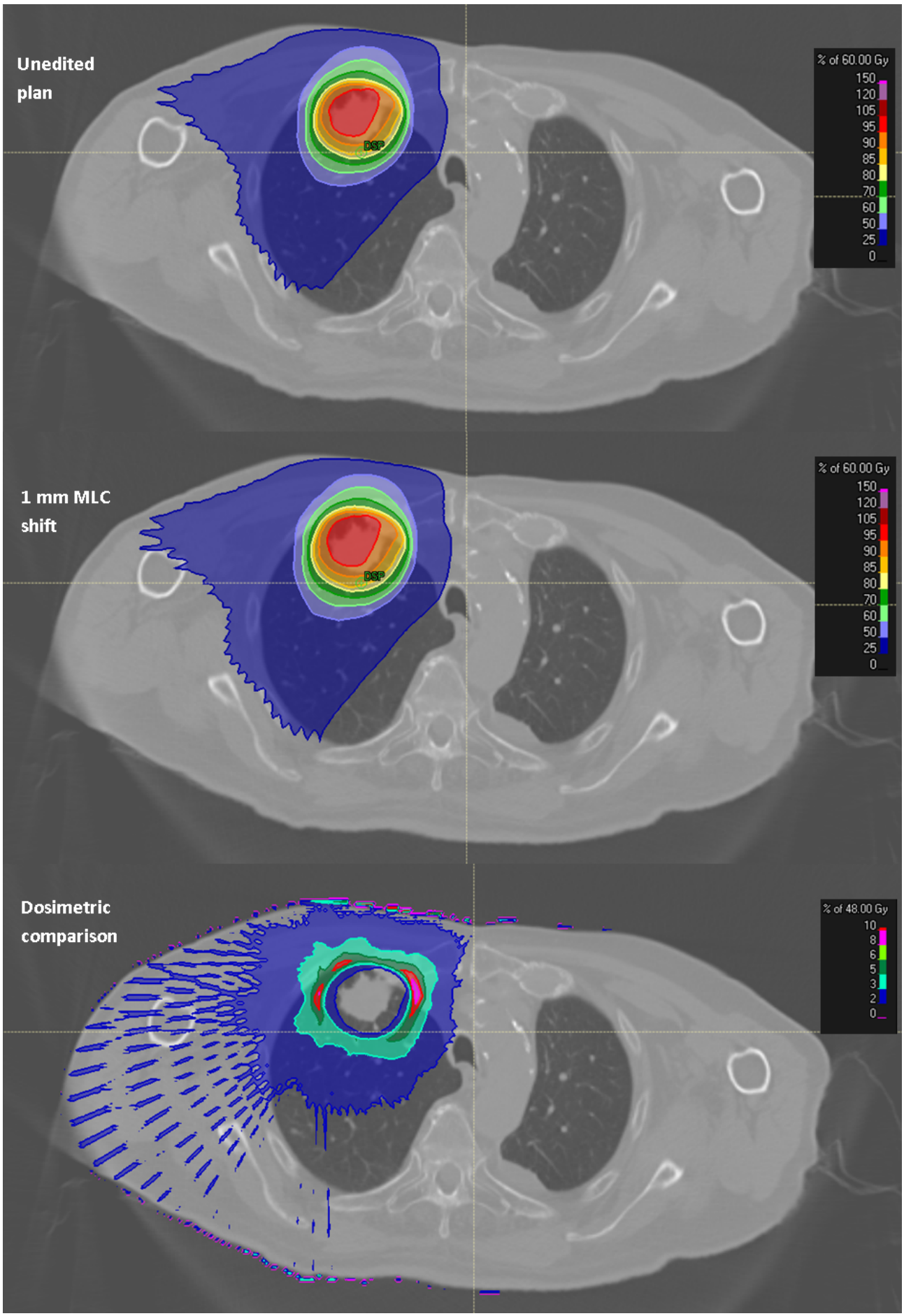


Figure 3.1: Transverse images of original plan, plan with 1 mm MLC errors introduced, and dosimetric comparison between two plans, respectively, for patient 3. The key for the dosimetric comparison shows the percent dosimetric difference between the two plans.

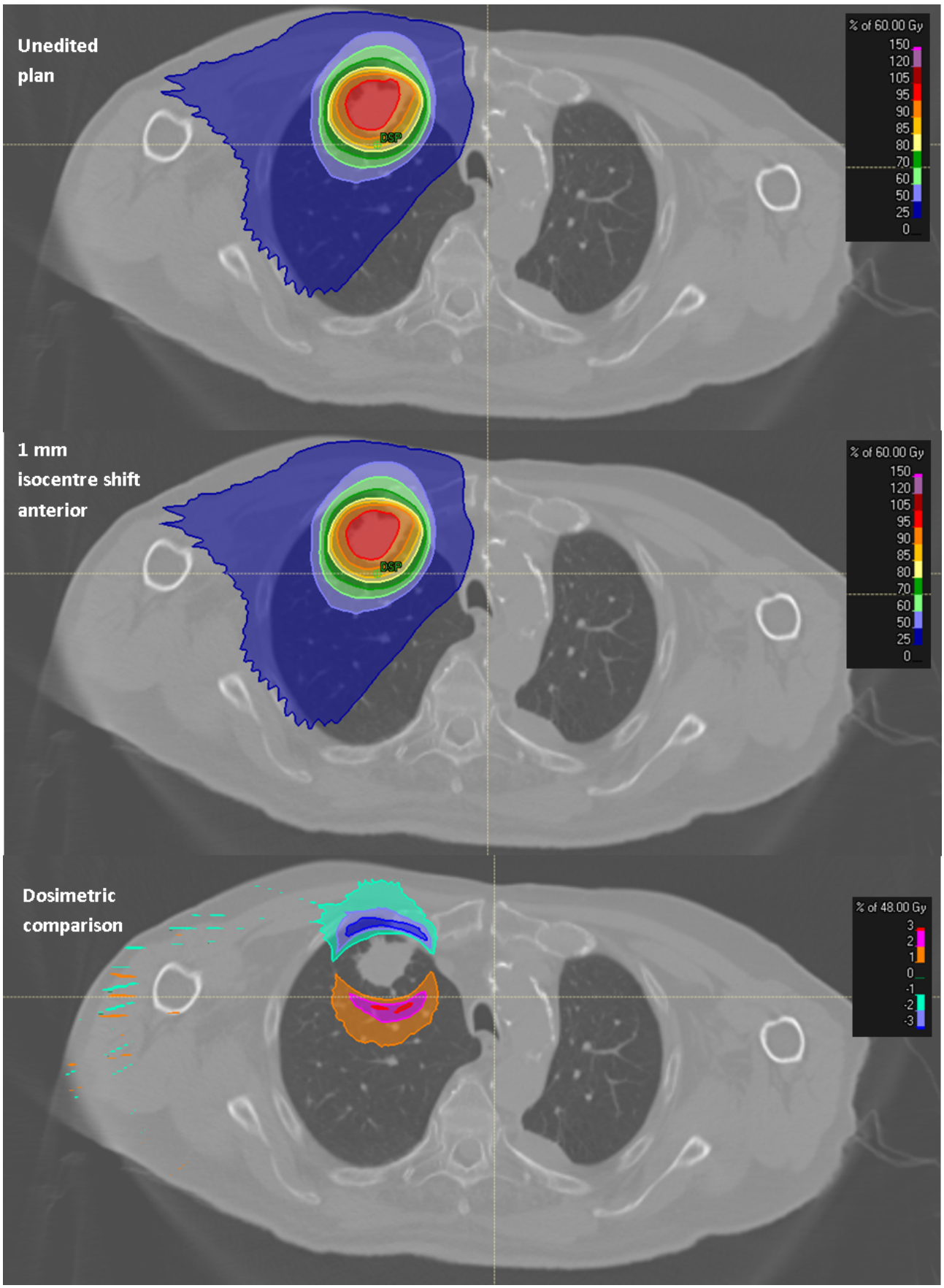


Figure 3.2: Transverse images of original plan, plan with 1 mm isocentre shift introduced, and dosimetric comparison between two plans, respectively, for patient 3. The key for the dosimetric comparison shows the percent dosimetric difference between the two plans.

approximately symmetrical on either side of the PTV. That is, an increase in dose on one side was matched in turn by an approximately equivalent decrease in dose on the other side of the target volume. This approximate symmetry was observed in most plans, although occasionally an asymmetry was observed for certain plans. This is likely due to the position of the PTV relative to the boundary of high and low density tissue, where the lung and ribcage interface, thus changing radiation scatter conditions for these regions.

3.1.2 DVH measurements

Maximum point dose measurements

Point dose measurements were performed for all patients. Figure 3.3 contains point dose measurements for patient 3, with and without introduced MLC class shift errors. Linear trend lines with associated uncertainties were drawn through the data, using least-squares regression in Excel. The standard error in these measurements was estimated at ± 8%; the reasoning behind this is discussed in Section 4.1.

Figures 3.5 to 3.8 contain the gradients produced from the trend lines in Figures 3.3 and 3.4. In these plots, the initial maximum point dose was measured from the original plan calculated in Raystation, while the gradient was measured from the linear trendlines plotted as a function of the introduced errors. As an example, in Figure 3.5, the point at (19.95, 1.45) refers to a single patient's maximum point dose calculated in the proximal bronchus. The introduction of MLC errors produced an increase in this maximum point dose, at a rate of 1.45 ± 0.12 Gy per mm (2 SD) of class shift error introduced to the plan. Hence if the plan was delivered with 1 mm MLC class shift error, one would expect this patient's ribs to receive a maximum point dose of 21.40 ± 0.12 Gy.

In theory, this is not a significant issue for this particular plan. The dashed lines in Figure 3.5 indicate the planning constraints for each OAR. For the proximal bronchus, the constraint is 30.00 Gy, meaning that the proximal bronchus for this patient received a tolerable dose. Assuming that the MLC position accuracy is kept within 1 mm (a parameter that is checked during routine linac QA), this OAR dose is only likely to exceed the constraint if a gross treatment error were to occur. Hence the introduction of MLC errors should not be of concern to the clinician for this particular OAR.

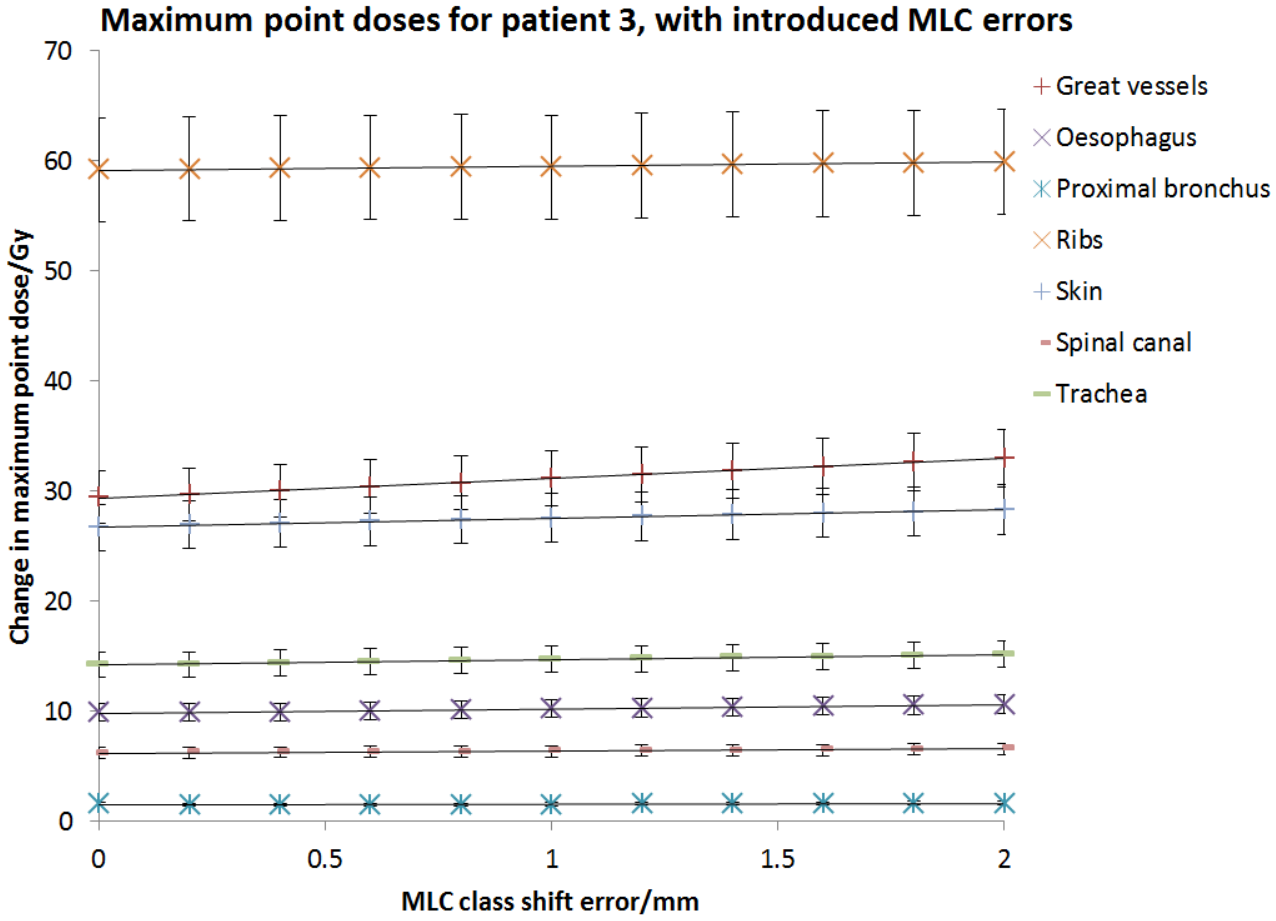


Figure 3.3: Maximum point dose measurements obtained for patient 3, for plans with and without introduced MLC class shift errors. For this patient, errors were introduced up to 2 mm, in order to ascertain the extent to which the trends obtained were linear. In the majority of patients, the gradients were calculated only up to 1 mm of introduced MLC error.

However, there are plans for which the introduced MLC error could have significant detriment to OARs. Take the point in Figure 3.5 at (38.56, 4.04), which refers to a single patient's maximum point dose to the ribs. The constraint for this OAR lies at 40.00 Gy; the plan is designed to achieve a maximum point dose to the ribs that is below this point. So the original plan has passed the constraint value for the ribs. However, it can be seen that the introduction of a class shift MLC error would eventually lead to this constraint being exceeded. Using the information provided, the expected MLC error required for failure can be calculated as:

$$Error = \frac{c - x}{y}$$

where c is the value of the constraint, in Gy, and (x, y) denotes the location of a point on Figure 3.5 (x being the initial maximum point dose, in Gy, and y being the gradient calculated from Figure X,

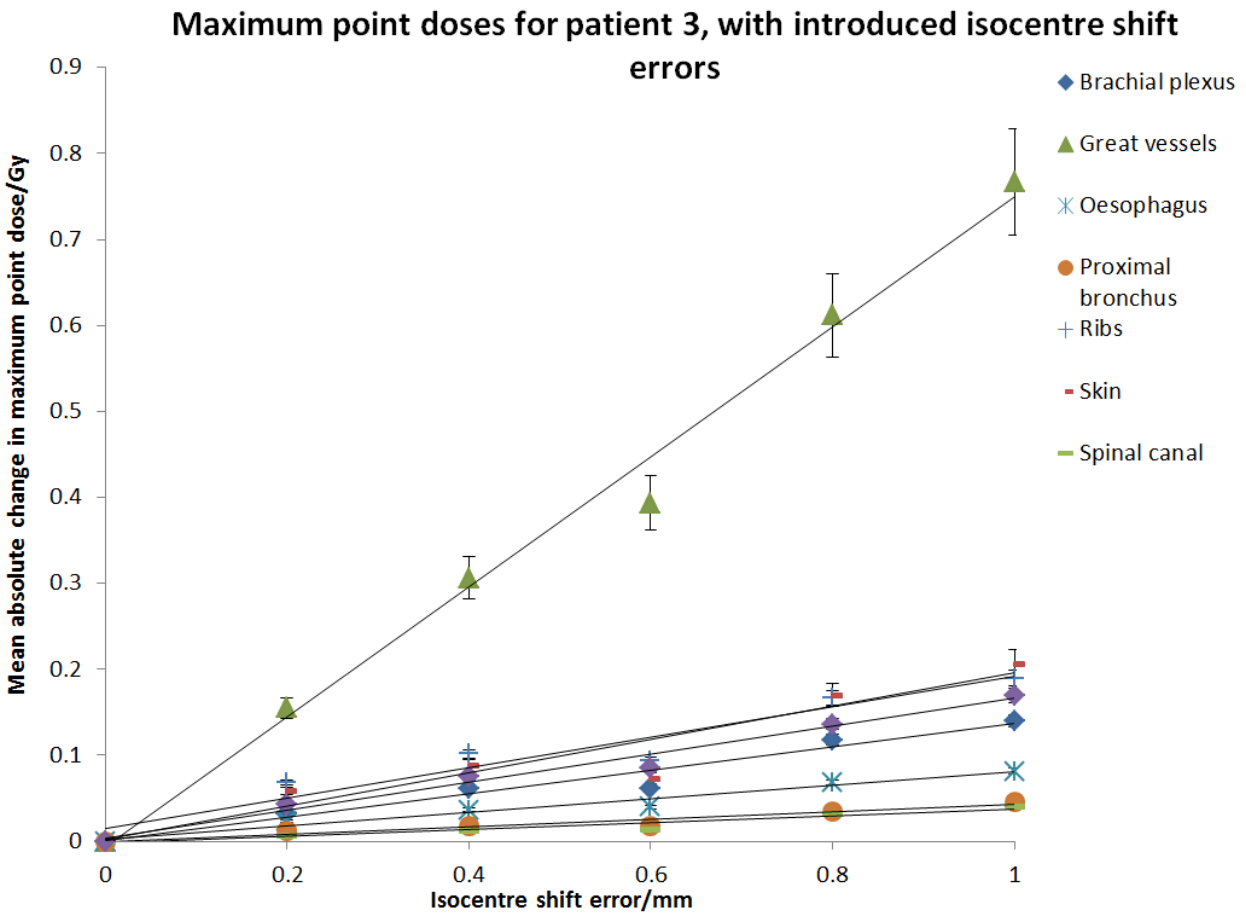


Figure 3.4: Maximum point dose measurements obtained for patient 3, for plans with and without introduced isocentre shift errors. The mean absolute point dose change was calculated based on the results from shifts in all six anatomical axes.

in Gy/mm). In the case shown above:

$$Error = \frac{40.00\,\text{Gy} - (38.56 \pm 3.08)\,\text{Gy}}{4.04\,\text{Gy/mm}} = 0.36 \pm 0.03\,\text{mm}$$

Hence one would only need to introduce an MLC class shift error of approximately 0.36 mm to exceed the constraint. In reality, it is unlikely that such an error would occur. MLC leaf errors tend to be random in nature, although systematic errors are often seen in individual leaves. The caveat of all this is that for certain plans, the introduction of errors will (in theory) have a minimal effect on the clinical effectiveness of the treatment plan. This is the case for plans where the DVH result for any specified metric isn't close to the tolerance for that metric. However, it is generally the case that there will be at least one metric that is close to tolerance, when compared to the standard uncertainty in the metric.

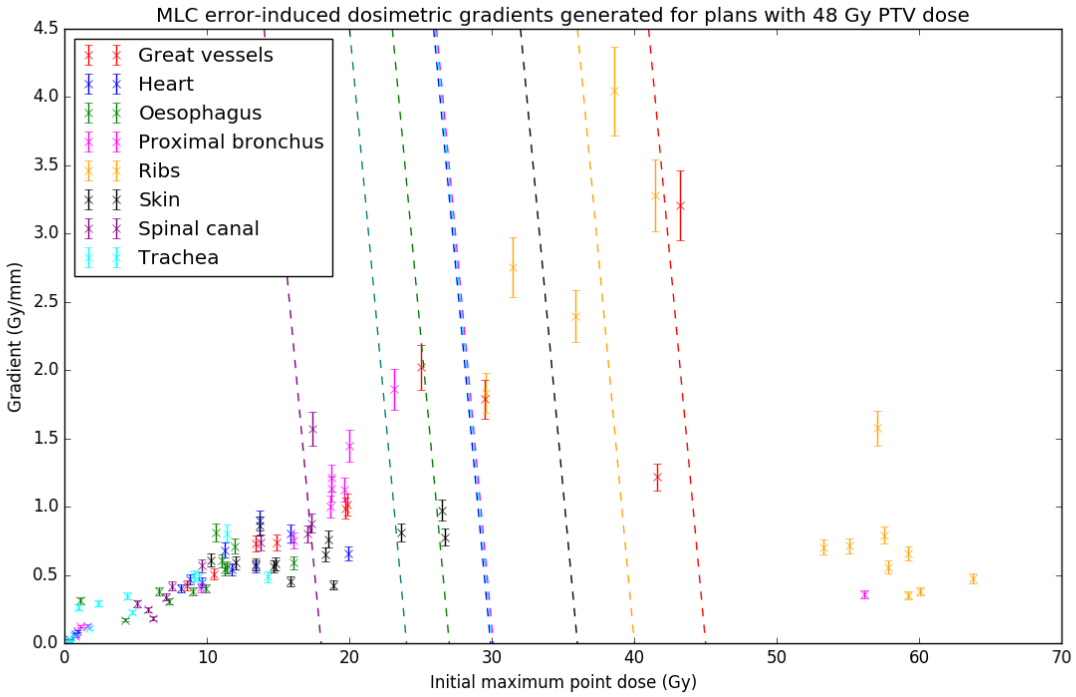


Figure 3.5: Maximum point dose gradients generated from introduction of MLC errors to plans with a PTV dose of 48 Gy. Dashed lines indicate thresholds at which associated OAR constraints should fail, if potential MLC-induced dosimetric error is taken into account. Error bars quoted to within ± 2 SD.

In almost all volumes calculated for each patients, gradients calculated for the isocentre shift trends were slightly affected by a systematic decrease in calculated dose, which appeared at the introduction of a 0.6 mm isocentre shift. This can be clearly seen in Figure 3.4, and can potentially be attributed to the dose grid, although the exact cause is uncertain. In general, the decrease in gradient that occurred as a result of including these points was less than 2% for all patients, which is negligible when considering the uncertainties involved in these measurements. Furthermore, the error appeared to be systematic for almost all volumes measured. Therefore, this phenomenon is unlikely to have had a significant effect on the trends observed in Figures 3.7 and 3.8.

Volumetric measurements

Two OARs in Table 2.1 had DVH constraints that analysed the percentage volume of the OAR receiving in excess of a threshold dose. For instance, the constraint for the whole lung in plans with 48 Gy prescribed to the PTV was $V_{5\ Gy} \leq 60\%$. The change in these values as a function of the introduced MLC and isocentre shift errors was also plotted, in a manner identical to that used for

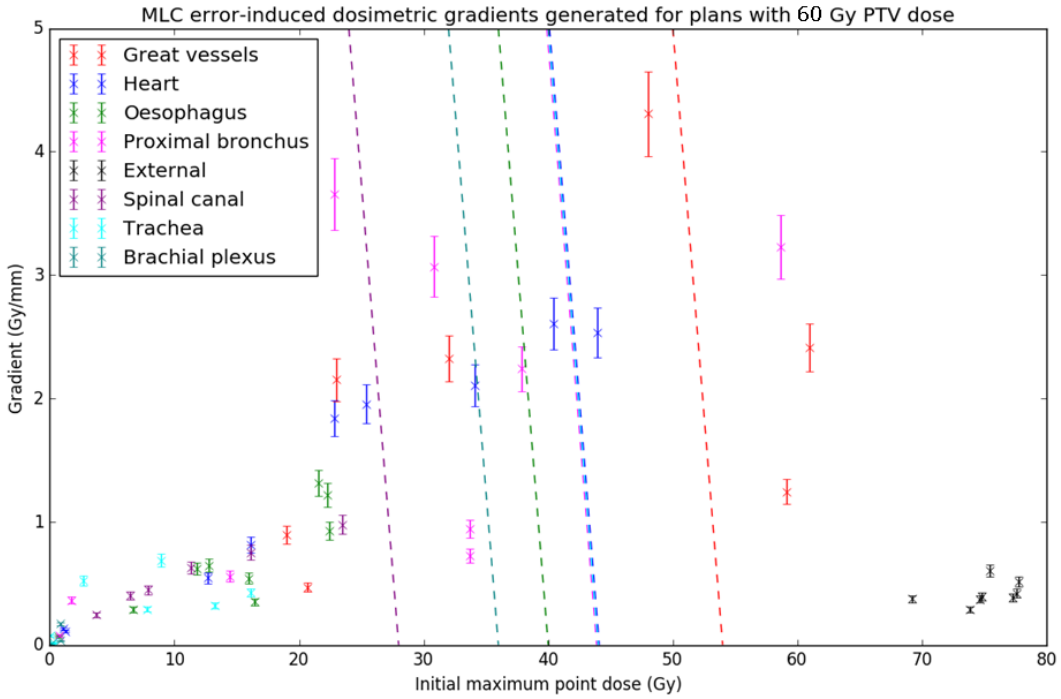


Figure 3.6: Maximum point dose gradients generated from introduction of MLC errors to plans with a PTV dose of 60 Gy. Dashed lines indicate thresholds at which associated OAR constraints should fail, if potential MLC-induced dosimetric error is taken into account. Error bars quoted to within ± 2 SD.

the maximum point doses. The gradients of the resulting graphs were calculated, and the results plotted. These results can be found in the Appendix (Figures 1 to 8). In general, the density of data acquired was too low to produce any quantifiable claims about trends. However, there were points that fell close to the boundary, in a similar manner to the maximum point dose measured in the ribs for one patient, referenced in Section 3.1.2.

3.2 Commissioning of MapPHAN

The majority of the commissioning process was concerned with the application of appropriate density overrides for the phantom, in order to correct for the angular dependencies of the diodes within the phantom. The results from this process are outlined in detail in the subsections below.

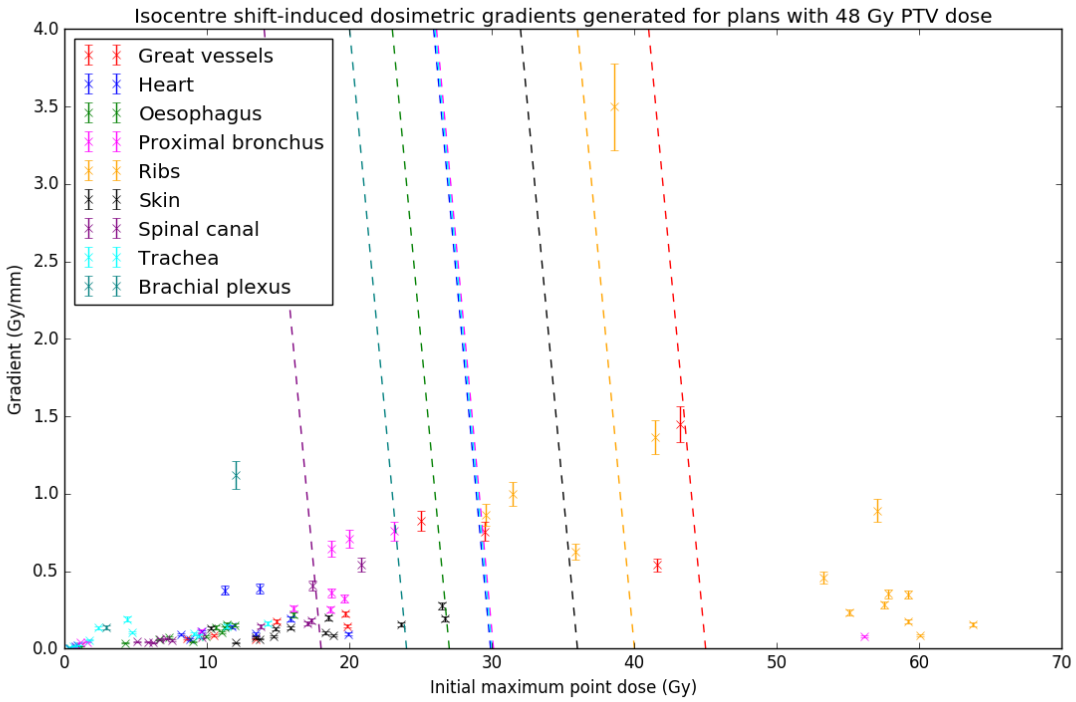


Figure 3.7: Maximum point dose gradients generated from introduction of isocentre errors to plans with a PTV dose of 48 Gy. Dashed lines indicate thresholds at which associated OAR constraints should fail, if potential MLC-induced dosimetric error is taken into account. Error bars quoted to within ± 2 SD.

3.2.1 Angular dependence of diodes and determination of density overrides

The process of determining appropriate density overrides for the phantom proved to be complex, as in addition to the inherent angular dependency of the diodes, several other confounding factors had to be accounted for. Sun Nuclear recommends that a uniform density override of 1.05 g cm^{-3} be used for the entirety of the phantom, including the diode array. However, the results indicated that this did not provide accurate agreement between measured and calculated doses delivered to the phantom.

Figure 3.9 shows the mean angular dependence of the central five diodes, with a uniform density override of 1.05 g cm^{-3} applied to the entire phantom in Raystation. The central five diodes are arranged in a 1 cm x 1 cm 'X' shape at the centre of the phantom. The maximum discrepancy between measured and calculated results was observed at gantry angles G90 and G270 (11.9% and 13.3% respectively). This is similar to results obtained in the literature; for instance, Jursinic *et al.*

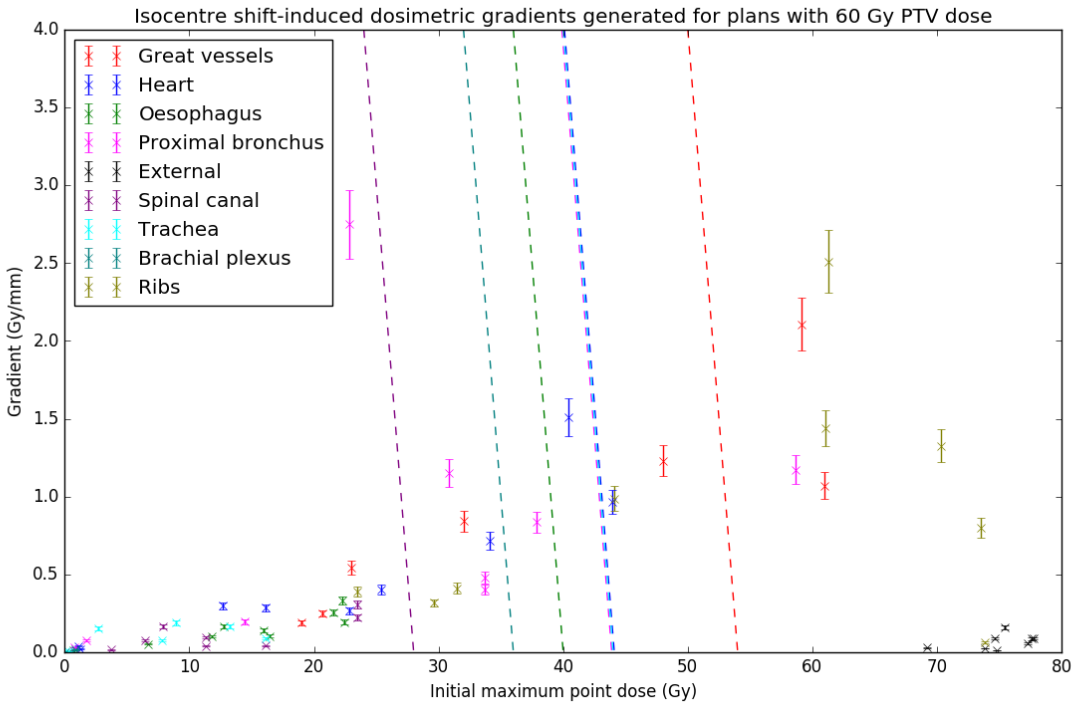


Figure 3.8: Maximum point dose gradients generated from introduction of isocentre errors to plans with a PTV dose of 60 Gy. Dashed lines indicate thresholds at which associated OAR constraints should fail, if potential MLC-induced dosimetric error is taken into account. Error bars quoted to within ± 2 SD.

found that application of a uniform density override generated a measurement over-response of 24% at G90 and G270, and an underresponse of 8% at G180, where the beam is delivered through the back end of the phantom [75]. Other centres have discovered similar results, but have overcome this by simply avoiding the delivery of beams at G90 and G270 [74]. The alternative suggested by Sun Nuclear is to rotate the phantom by 90° into the sagittal orientation, for beams that are delivered from these directions. Since this technique does not work for most rotational beam deliveries, which typically subtend a wide arc of gantry angles, it was decided to reduce the angular uncertainty to within ± 2% if possible via the application of density overrides in the TPS.

Optimization of density overrides

Figure 3.10 shows the final results for the angular dependence obtained using the optimized density overrides found in Figure 2.3. The data shown here represents the measurements for diodes within ± 40 mm laterally of the central MapCHECK diode. Results for all diodes can be found in Figure 9 in

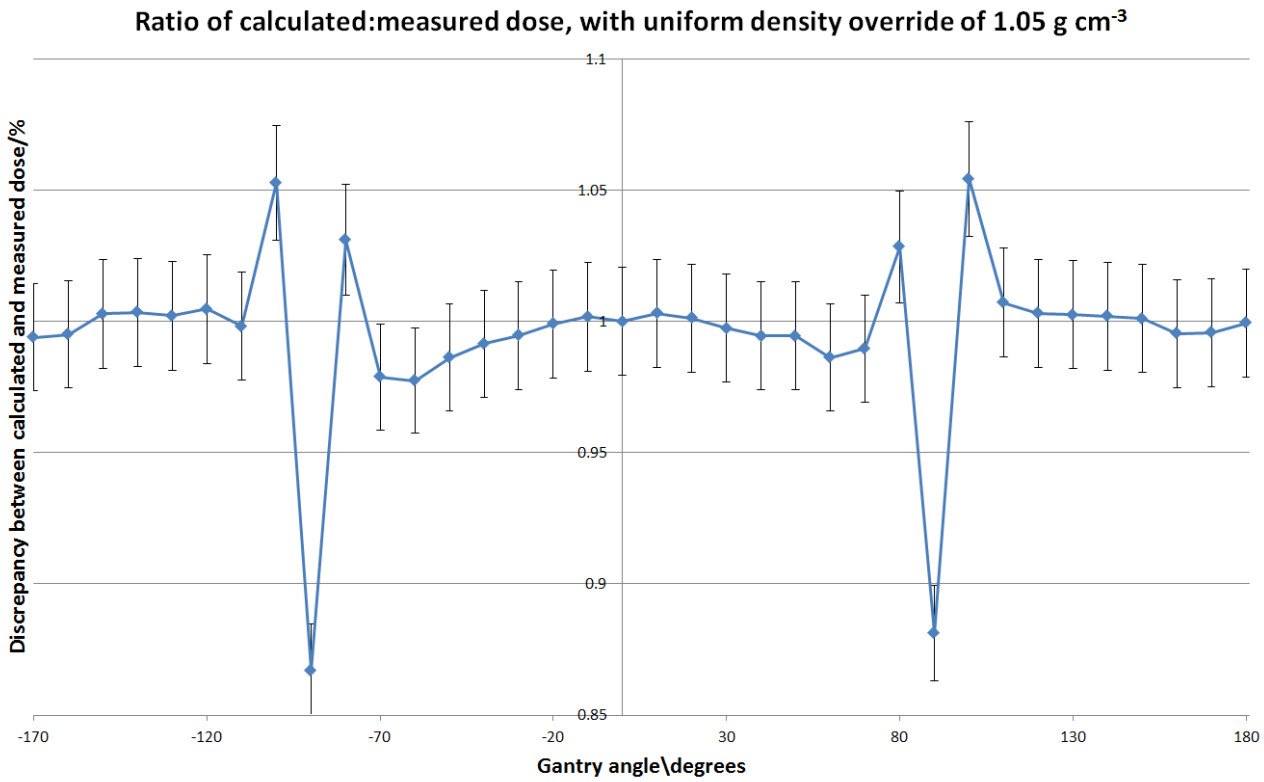


Figure 3.9: Mean ratio of calculated/measured dose for central five diodes of MapCHECK in MapPHAN, with uniform density override of 1.05 g cm^{-3} applied to entire phantom. Error bars represent uncertainty of ± 2.06%, which combines in quadrature the uncertainty of central beam axis dosimetry in Raystation with the relative dose measurement uncertainty of MapCHECK. All errors quoted to within ± 2 SD.

the Appendix. The inclusion of measurements from diodes at a distance greater than 40 mm from the central diode significantly increases the variation in dosimetric discrepancy between measured and calculated doses for individual diodes. The maximum angular dependency of the central diodes was found to be equal to ± 4.15% for the coronal orientation, and ± 3.20% for the sagittal orientation. The assumption was made that there was no significant aliasing in these measurements, although it is possible that the angular dependency is higher than the figures stated here.

Sun Nuclear recommends that the electron density applied to the phantom should be equal to that of water. During the process of modifying the density overrides, it was found that changing the electron density had a negligible effect on dosimetry when compared to changing the physical density. This was also the case for the diode array itself, which contains materials with markedly differing electron density to that of water. However, reference to Figure 1.3 indicates that for 6 MV X-ray beams, pair production becomes more predominant for higher atomic number materials, such as those found in the array. This has the potential to significantly affect the Raystation-calculated

dose in these regions, especially for lateral beams passing parallel to the array. Despite this, it was decided to maintain a uniform electron density throughout the entire phantom, since the problem of optimizing the angular dependency was already a complex multi-variate problem. The accuracy of the results obtained was sufficient to back up this methodology.

The MapCHECK phantom itself consists of multiple layers of plastic, which were not identical in density and had to have density overrides applied to reflect this. Additionally, there was found to be a small cylindrical hole in the posterior plastic backing of the phantom, directly below the central diode. Hence adding this feature to the model significantly improved dosimetric results for the central diodes, which were directly adjacent to this hole.

Since lung SABR plans have such a high dose falloff, the majority of diodes outside of this central diode region will register doses lower than 10% of D_{max}, meaning they are not considered for the gamma calculation. For certain plans, a very low number of diodes outside of the central diode region exceeded the gamma threshold while simultaneously exceeding the 10% threshold required for inclusion in the measurement. An example of these points can be seen in Figure 3.13. The gamma comparison shows some points well outside of the central PTV dose region which are nonetheless exceeding the gamma threshold comparison. These points are closest to the left edge of the phantom (which is set up in the coronal orientation). For this plan, the range of gantry angles utilised was from G181 clockwise through to G20, meaning that the majority of radiation was delivered passing through the left side of the phantom. The resultant elevation in dose on this side of the diode array allows these points to exceed the 10% threshold required for the points to be included in the gamma analysis.

These diodes are theoretically beyond the buildup region for 6 MV X-rays in terms of water-equivalent thickness of material, so the reason for this discrepancy is not immediately clear. The effect could potentially be explained by the fact that closer to the edge of the phantom, the density override for the diode array has a greater effect on the Raystation calculation. This conclusion can be reached from a geometric basis. The arc subtended by the beam as it passes through the phantom to reach the diodes at this point will pass laterally through the phantom for a greater proportion of the arc than it would for diodes at the centre of the phantom. It is possible that this phenomenon had a biasing effect on the overall gamma results, since the homogeneity and baseline of the low dose region would have varied between plans as a function of the gantry angles

used for beam delivery. As a result, this effect was more pronounced for certain plans where a higher proportion of dose was delivered laterally through the array. Overall, the gamma pass rate did not decrease too significantly for these plans - the case shown in Figure 3.13 was one of the worst examples, and still produced clinically acceptable gamma pass rates. However, it is certainly something to keep in mind when performing QA, as this effect could generate a number of false positives if it leads to QA failures.

Comparison between coronal and sagittal orientations

The mean dose discrepancy for the central diodes is shown in the trend lines of Figure 3.10. Overall, the measured dose was slightly higher than the calculated dose for both the coronal and sagittal orientations. As the density overrides were chosen based on the angular dependency results in the coronal orientation, the mean dosimetric discrepancy between measured and calculated doses for these angular dependency beams was slightly lower than for the beams delivered to the phantom in the sagittal orientation, which had identical density overrides applied. While the discrepancy was still within the recommended tolerance of ± 2%, it is significant enough to be taken into account when performing PSQA.

The mean discrepancy for the central diodes across all gantry angles is 0.11% greater for the sagittal orientation than for the coronal orientation (0.39% and 0.28% respectively). While this is a small discrepancy overall, the number of diodes from which these statistics were measured is great enough to suggest that this difference is significant, and attributable to something other than output fluctuations in the linac. While the phantom is rotated, in theory this should have no effect on the dose it measures, as the linac is designed to produce consistent radiation output independent of gantry angle. Ignoring phantom rotation, the other major change in experimental setup is the position of the couch relative to the central diodes. With the phantom positioned at isocentre, the couch is closer to the average diode position when the phantom is in the coronal orientation. It can be inferred that dosimetric uncertainties propagated into the angular dependency measurements that are inherent to the couch model will have a greater effect on results obtained with the phantom in the coronal orientation. This is because more beams passed through the couch to reach the phantom when it was in the coronal orientation, as a consequence of the couch being moved closer to isocentre while positioning the phantom. While this is a possible cause of the 0.11% discrepancy, further

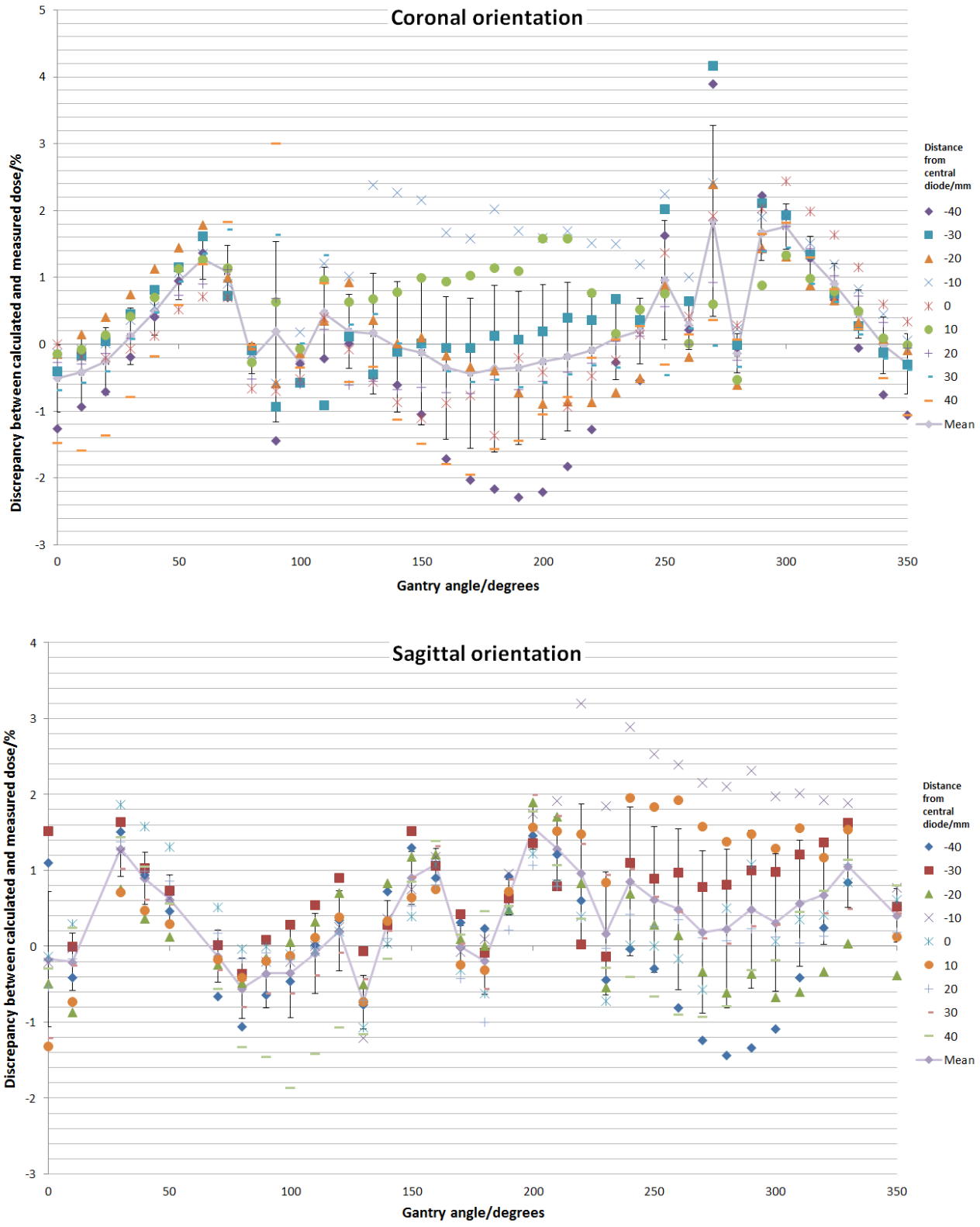


Figure 3.10: Ratio of calculated/measured dose for diodes within ± 40 mm of central diode along central transverse axis, for angular dependency measurements on MapCHECK in coronal (top) and sagittal (bottom) orientations. Trend line represents mean discrepancy for each gantry angle. The mean discrepancy for the central diodes across all gantry angles was 0.28% for the coronal orientation, and 0.39% for the sagittal orientation. Error bars represent 1 SD from the mean discrepancy for all central diodes.

measurements could be performed to determine the true extent of the couch model uncertainty. The couch model was initially commissioned in such a way that 1% dosimetric accuracy was attainable in the TPS, so this error is certainly comparable to any uncertainty in the couch model's dosimetric accuracy. Unfortunately, the commissioning data for the couch model could not be obtained at the time of writing.

Standard deviation of dosimetric discrepancy as function of gantry angle

SNC Patient software provides a histogram of measured versus calculated values when gamma analysis is performed. From this, the mean disparity for diodes along the central transverse axis of the phantom was calculated, using the following equation:

$$D_{mean} = \frac{\sum_i (f_i \times D_i)}{\sum_i (f_i)} \tag{3.1}$$

where D_i = (measured dose - calculated dose)/(calculated dose) for histogram bin i, and f_i is the number of measurements for this bin. The standard deviation of the disparities for all central axis diodes was calculated via the following equation:

$$\sigma = \sqrt{\frac{\sum_i (f_i \times (D_i - D_{mean})^2)}{\sum_i (f_i) - 1}} \tag{3.2}$$

The final density chosen for the array itself was 0.80 g cm^{-3}, going counter to the idea that the array consists of high-Z materials, which typically have a greater density than that of water (1 g cm^{-3}). However, it can be seen upon close inspection of the CT scan that the array consists of both circuitry and a large number of air pockets, which significantly reduced the mean density of the array as a whole.

In order to compensate for the inherent angular dependency of the diodes, the density override for the array was altered so that the Raystation-calculated dose distributions for beams passing parallel to the plane of the array were optimally aligned with the measured doses of the most central diodes to within the desired tolerance of ± 2%. Figure 3.11 shows that applying this density override provided a central region of approximately 80 mm width that was acceptable for use in PSQA of arc

treatments, based on the 2% tolerance. The extent of this agreement can be used to determine the restrictions that should be placed on the device when using it for PSQA - i.e. limitations on the PTV size that can be accurately measured with gamma analysis on the device.

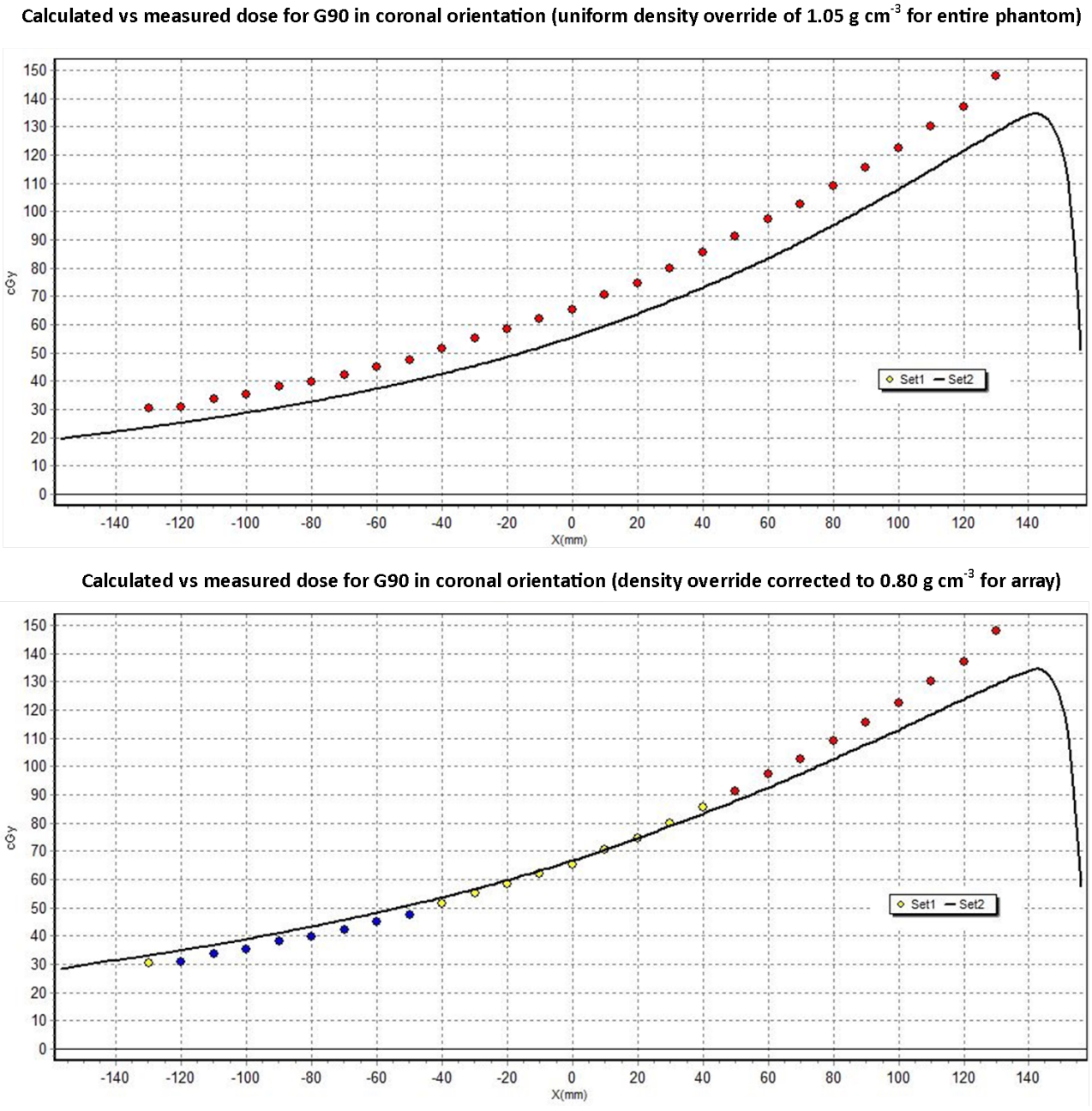


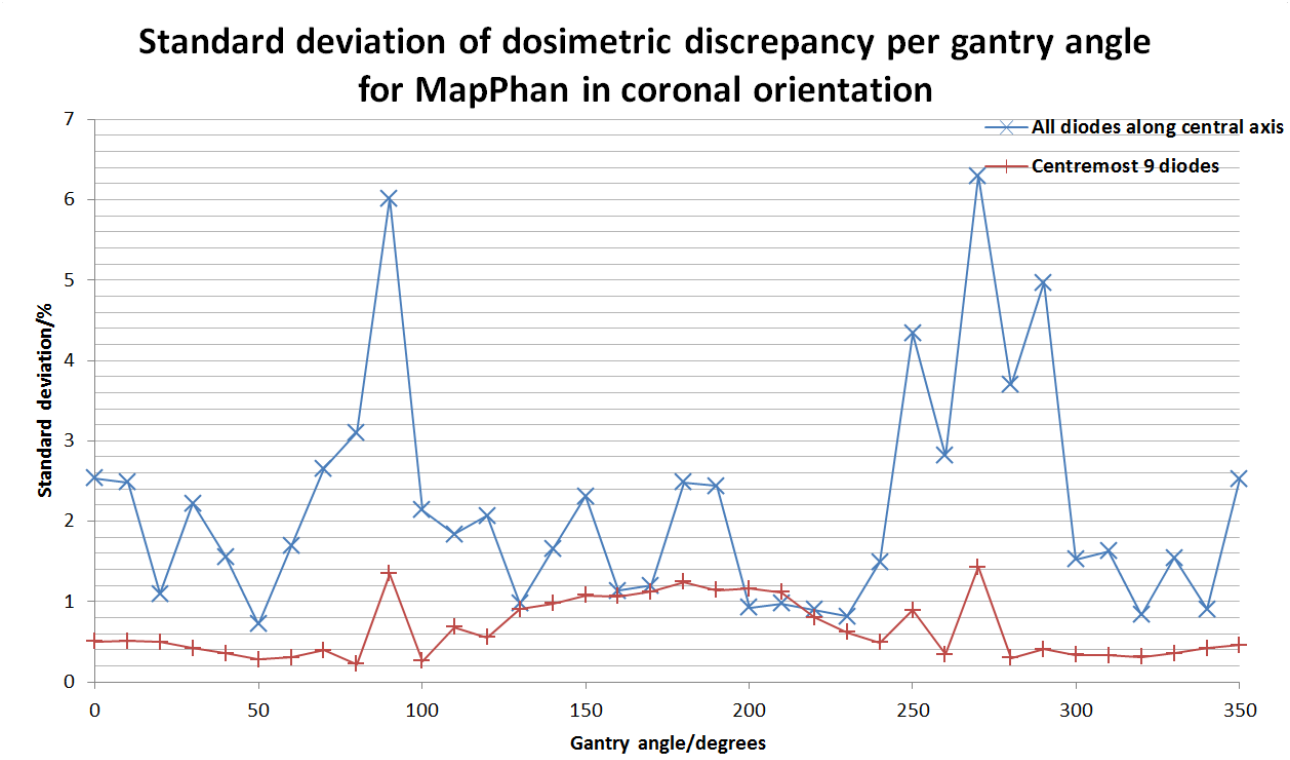
Figure 3.11: Screenshots of agreement between the measured dose for MapCHECK 2 in coronal orientation (Set 1, with individual points representing the dose measured at each diode at distance X mm from the central diode in the transverse plane), versus Raystation calculated dose (Set 2, black line). The upper figure shows the agreement when a uniform density override of 1.05 g cm^{-3} was applied to the entire phantom. The lower figure shows the agreement following the correction of the array override to 0.80 g cm^{-3}. Yellow points indicate agreement within ± 2% between the measured and calculated dose, with red and blue points representing diode measurements that lay outside of this tolerance.

Figure 3.12 shows the SD calculated for the diodes along the central axis of the phantom, for both coronal and sagittal orientations. Compared to the calculation of mean discrepancy, the SD provides additional information about the degree to which outliers are present in the dataset. Hence a large standard deviation with a low mean discrepancy likely indicates that there is an offset in the discrepancy that is balanced out on either side of the array, as can be seen in Figure 3.11.

It should be noted that a larger discrepancy is expected when the peripheral diodes are included, as the 10 x 10 cm^2 fields used in these measurements did not cover the entire array. Hence some of the diodes will only be receiving scattered and leakage radiation, primarily from the phantom and linac head. Dosimetric accuracy in these out-of-field regions is expected to have greater uncertainty. However, beams delivered at gantry angles within 10^o of the array plane (G90/G270 for coronal, and G0/G180 for sagittal orientation) included all central axis diodes in the central beam region. (Even beams delivered at gantry angles 20^o from the array plane, such as G70 for the coronal phantom, included over 90% of the central axis diodes in the central beam region.) Despite this, the SD for all diodes is highest in these regions where the beam is delivered parallel (or close to parallel) to the array plane. This is further evidence of the inherent angular dependencies of the diodes themselves.

An additional complication is the increase in SD observable for beams delivered through the back end of the phantom. For the coronal orientation, this is seen between approximately G110 and G250, whereas for the sagittal orientation, this is observed between approximately G200 and G340. Sun Nuclear have stated that the diodes in MapCHECK were primarily designed for radiation delivered through the front end of the phantom, implying that the dosimetric uncertainty for beams delivered through the back end of the phantom is higher. There is also a greater water-equivalent thickness of material for the beam to penetrate to reach the diodes through the back end of the phantom, and different material overrides were applied to the top and bottom layers of the MapCHECK phantom. Both of these factors may be contributing to the increase in SD.

It was concluded that utilising MapCHECK in MapPHAN for the QA of conformal arc SABR plans is clinically acceptable, as long as appropriate restrictions are set in place for use of the device. The choice to limit the device's bounds to $\pm$ 4 cm from the central diode column is conservative, based on the misaligned dose distributions seen in Figure 3.11. However, this sets approximate boundaries for consistent 2% dosimetric agreement with Raystation, entirely independent of the angle of radiation delivery.



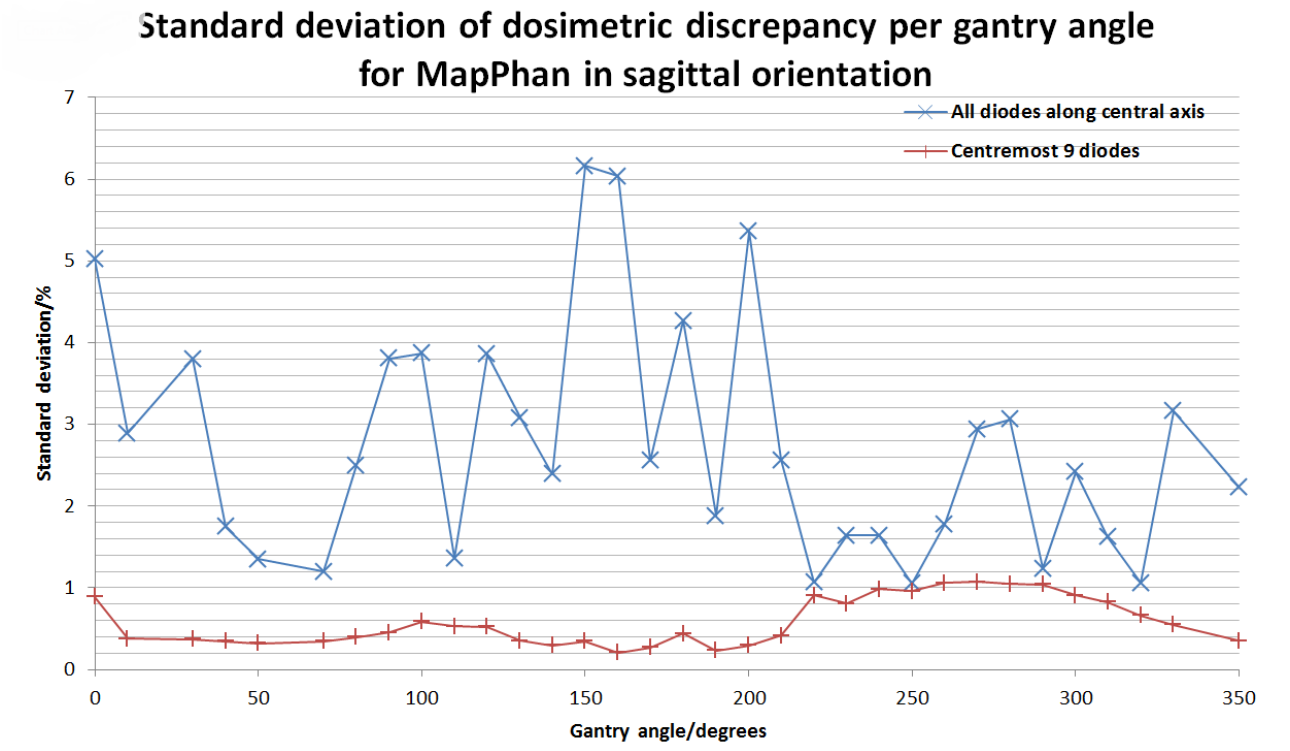


Figure 3.12: Standard deviation as a function of gantry angle for MapPHAN in both coronal and sagittal orientations.

3.3 Patient specific quality assurance measurements

3.3.1 Conformal arc commissioning measurements

In order to confirm the validity of the density overrides in the previous section, three conformal arc beams were delivered to MapCHECK in MapPHAN, using a 10 x 2 cm^2 field, a 10 x 5 cm^2 field, and a 10 x 10 cm^2 field. The measured results were compared to their respective calculated dose distributions via gamma analysis, with 1%/1 mm criteria. The resultant gamma pass rates were 95.2%, 95.3%, and 83.9% respectively, for the field sizes listed. The measured and calculated dose at the point of the central diode was also recorded for all three measurements, with the disparity being equal to -2.22%, -1.59%, and -1.29% respectively. This indicated that the calculated dose was systematically higher than the measured dose, with the disparity increasing for smaller field sizes. It was decided that these results were acceptable for commissioning purposes, given that the tolerance for 2%/2 mm gamma criteria at ADHB is 85%, and the gamma criteria being used here were more strict than this. Additionally, the tolerance for absolute dose disparity is ± 3%, and all central diode disparities were less than this value. This test was not an entirely fair representation of the plans that were actually delivered, since it utilised the linac jaws to set field sizes rather than the MLC, which has a greater transmission factor for radiation. However, the difference between transmission factors was negligible compared to the 3% absolute dose tolerance required. Hence it was decided to continue with measurement of actual patient plans.

3.3.2 PSQA measurements

Nine patients with plans that had introduced errors were delivered to the phantom - in total, 198 plans were generated for the final analysis. All plans were delivered on the same Varian Clinac iX System linear accelerator (Varian Medical Systems, Palo Alto, California). SNC Patient software was used in conjunction with MapCHECK for these measurements. Two measurements were taken for each plan, with the phantom offset by 5 mm superior via couch shift for the second measurement. These two measurements were then merged in SNC Patient software, thus doubling the effective dosimeter resolution of the phantom.

It was assumed that the linac used for measurements was able to deliver both the modified and

error-free plans with equal bias. For example, any MLC delivery error was assumed to be consistent between these two groups, regardless of the degree to which test errors were introduced. This assumption applied to all variables on the linac (e.g. collimator angle, gantry angle, MU output, etc.). In reality, an accumulation of these random and systematic errors will have reduced the accuracy of gamma analysis for all plans.

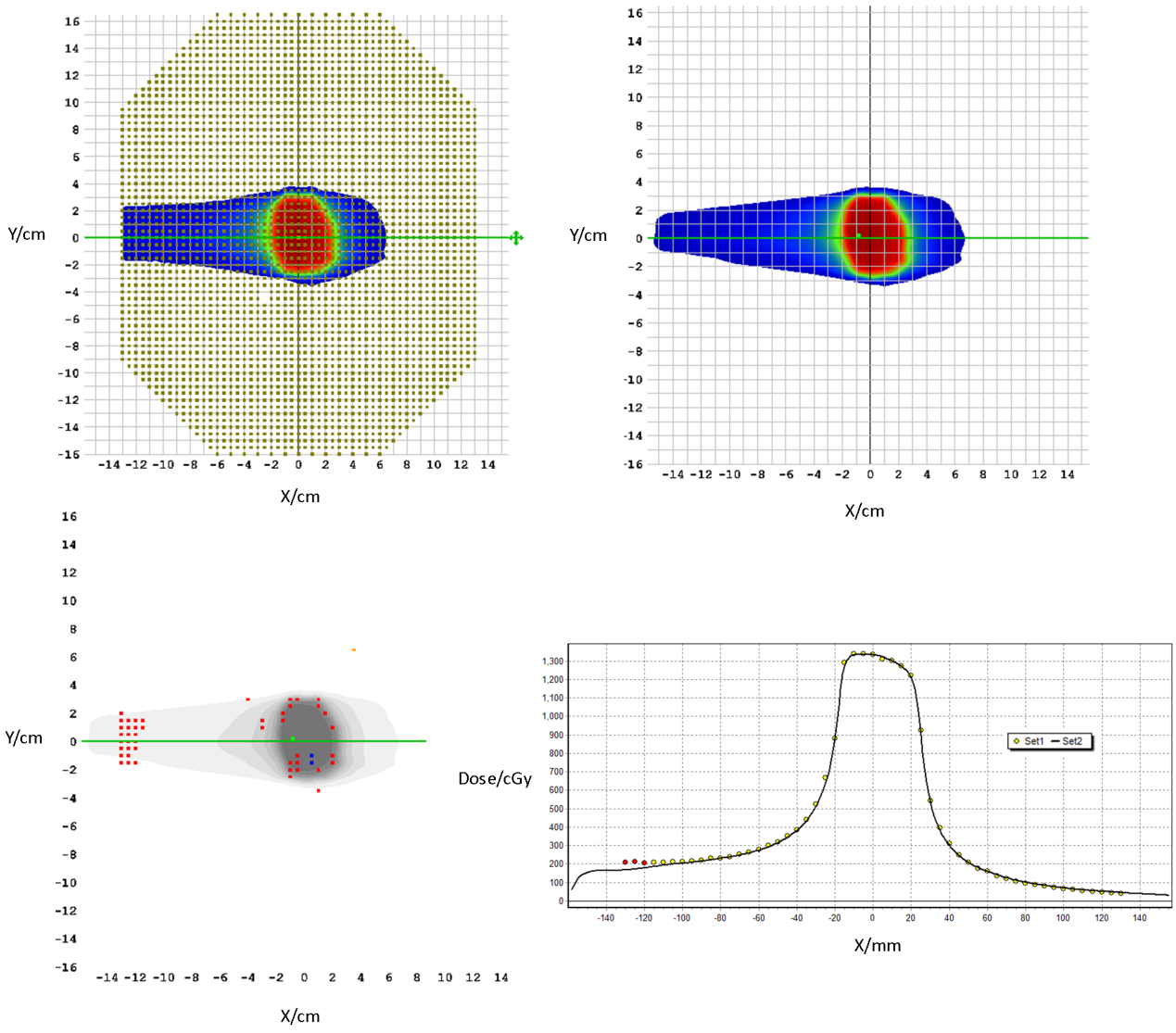


Figure 3.13: Example of gamma analysis for one patient, using 1%/1mm gamma analysis criteria. Top left: measured doses as delivered to MapCHECK in MapPHAN. Top right: Raystation calculated doses. Bottom left: agreement between measured and calculated doses for individual diodes on MapCHECK. Red points indicate that the dose received by MapCHECK was greater than the Raystation calculated dose at that point (outside of the 1%/1mm threshold), while blue points indicate that the measured dose was lower than the calculated dose. Bottom right: dose profile through MapCHECK at the level of the green line.

3.3.3 ArcCHECK measurements

Film measurements

As a 'gold standard' for reference, film measurements were taken to confirm the validity of the plans being delivered, with film inserted into ArcCHECK in both coronal and sagittal orientations. The high resolution capabilities of film were used to confirm that MapCHECK in MapPHAN would be sufficient for analysis of the high dose gradient regions present in the SABR plans being delivered.

Table 3.1 contains the results for relative film measurements performed on both plans calculated in Raystation and Pinnacle. Three of these measurements fell just outside of the 2%/1 mm threshold of 85% passing points, with the worst of these being at 84.1%. This indicated that there was generally good agreement between the treatment planning systems and the delivered doses. In addition, three of the plans had a disparity of greater than 5% between the Pinnacle-calculated dose distribution and the Raystation-calculated distribution, for their respective gamma passing rates. For two of these plans, using Raystation to calculate the dose actually increased the gamma pass rate, indicating that it would be acceptable to use Raystation for calculation of conformal arc plans in place of Pinnacle, where the plans were originally calculated.

Table 3.1: Gamma analysis results comparing Raystation- and Pinnacle-calculated doses to measured doses for film measurements taken in ArcCHECK phantom, measured in both coronal and sagittal orientations. Three original patient plans were measured. The thresholds for passing the analysis were 85% for 2%/1 mm and 90% for 3%/1 mm.

		1		2		3	
		Coronal	Sagittal	Coronal	Sagittal	Coronal	Sagittal
Raystation	2%/1 mm	88.0	95.6	84.1	90.1	89.3	85.5
	3%/1 mm	94.4	99.7	94.2	98.7	98.9	98.9
Pinnacle	2%/1 mm	87.3	84.2	95.6	84.3	85.2	86.3
	3%/1 mm	93.7	95.9	97.8	94.0	97.0	98.5

Point dose measurements

As a final, gold-standard validation of the accuracy of delivery of the original plans, point dose measurements taken on the nine original plans were analysed. These plans were delivered to the ArcCHECK phantom with acrylic insert, using a 0.04 cm^3 CC04 ionization chamber placed at its

centre to measure the dose. The couch was shifted for each plan to ensure that the chamber was positioned in a high-dose, shallow dose-gradient region, which simultaneously increased statistical counting accuracy and decreased geometrical uncertainty. The recorded dose for each plan was normalised to the dose received by the chamber after delivery of 200 MU of 6 MV X-rays through a 10 x 10 cm^2 field at G0. This normalised dose was then compared to the normalised doses calculated in Raystation for the same phantom. The expected agreement between the measured and calculated doses was ± 3%, as per recommendations laid out by IAEA Human Health Series No. 31, ESTRO Booklet No. 9 and similar publications [83, 84].

The agreement between measured and calculated dose was well within the requisite tolerance for all 9 plans, as shown in Table 3.2. This, coupled with the agreement between the measured and calculated central diode readings on MapCHECK in MapPHAN, along with the gamma pass results, provided sufficient evidence that the phantom was ready for clinical use.

Table 3.2: Point dose measurements recorded with CC04 ionization chamber placed inside ArcCHECK phantom. All plans fell within the specified tolerance of ± 3%.

Patient number	Calculated dose/Gy	Normalised measured dose/Gy	Agreement/%
1	7.06	7.04	0.3
2	11.99	11.90	0.7
3	12.56	12.53	0.2
5	14.00	14.06	0.4
6	6.51	6.44	1.1
7	6.53	6.52	0.1
8	11.07	11.06	0.0
9	12.79	12.80	0.1
10	11.04	11.10	0.5

3.4 Analysis of MapPHAN measurements

Table 3.3 contains the mean gamma pass rates for plans with and without introduced errors, for all gamma thresholds used. The mean disparity between the measured and calculated dose at the central diode is also included.

Table 3.3: Mean gamma pass rates for plans with and without introduced errors, for varying gamma thresholds. Mean disparity between measured and calculated dose at the point of the central diode was also tabulated for the merged results, with shifts included in SNC. Errors were calculated at 1 S.D. based on the population for each result.

		Gamma pass rate for merged results/%						**Gamma pass rate for unmerged results/%**					
		1%/1mm	*2%/1mm*	*2%/2mm*	*3%/2mm*	*3%/3mm*	*5%/1mm*	*1%/1mm*	*2%/1mm*	*2%/2mm*	*3%/2mm*	*3%/3mm*	*5%/1mm*
Introduced isocentre shift for coronal orientation	*0*	87.5±3.8	93.9±3.3	98.0±2.1	99.1±1.1	99.5±0.8	98.1±0.8	87.4±4.3	94.1±3.5	97.8±2.1	99.0±1.0	99.6±0.7	98.1±1.1
	1 mm left	84.9±5.9	92.4±4.1	97.7±1.5	99.1±0.7	99.5±0.6	97.4±1.4	85.1±5.9	92.7±3.8	97.7±1.4	98.9±0.7	99.5±0.6	97.4±1.5
	1 mm right	85.9±4.6	93.1±3.1	97.7±1.7	98.9±1.0	99.4±0.7	98.3±1.1	85.8±4.1	93.3±3.4	97.7±1.6	98.9±1.0	99.5±0.7	98.7±0.9
	1 mm sup	85.2±5.0	92.5±3.2	97.2±1.9	98.8±1.0	99.3±0.8	97.7±1.2	85.4±5.3	92.8±3.5	97.4±1.9	98.7±0.9	99.3±0.7	97.9±1.7
	1 mm inf	84.3±4.7	92.1±2.8	97.3±1.7	98.8±0.9	99.3±0.8	97.5±1.2	85.1±4.5	92.6±3.0	97.2±1.8	98.8±1.0	99.5±0.7	97.9±1.4
	1 mm post	84.7±5.6	92.1±3.7	96.9±2.3	98.3±1.4	99.3±0.8	97.2±1.7	84.3±5.6	91.8±4.0	96.9±2.5	98.3±1.4	99.4±0.8	97.3±1.8
	1 mm ant	86.4±6.1	93.6±3.4	97.9±1.5	99.0±0.9	99.4±0.7	98.2±1.3	86.0±5.6	93.5±3.6	97.9±1.5	99.0±0.9	99.5±0.6	98.1±1.4
Introduced isocentre shift for sagittal orientation	*0*	88.5±2.6	95.4±2.1	98.9±0.6	99.8±0.2	99.9±0.1	98.8±1.2	88.8±3.3	95.6±2.5	98.9±0.8	99.9±0.2	100.0±0.1	98.9±1.3
	1 mm left	88.1±4.6	95.6±2.2	98.7±0.8	99.6±0.5	99.9±0.2	99.1±0.8	88.1±5.0	95.7±2.2	98.8±0.9	99.7±0.5	100.0±0.1	98.9±1.1
	1 mm right	88.4±4.4	95.4±2.5	98.5±1.1	99.6±0.4	99.9±0.1	98.9±0.6	88.7±4.4	95.4±2.7	98.4±1.3	99.6±0.4	100.0±0.1	98.9±0.8
	1 mm sup	90.3±2.7	96.8±1.1	99.0±0.6	99.8±0.1	100.0±0.1	99.6±0.3	90.4±3.1	97.2±1.5	99.2±0.9	99.9±0.2	100.0±0.1	99.6±0.5
	1 mm inf	90.3±3.9	96.8±1.0	99.0±0.5	99.9±0.1	99.9±0.1	99.7±0.4	90.4±3.7	97.2±1.4	99.2±0.8	100.0±0.1	100.0±0.0	99.6±0.4
	1 mm post	89.7±3.6	96.5±1.6	98.9±0.6	99.9±0.1	100.0±0.1	99.6±0.3	89.4±4.2	96.6±1.9	99.0±0.9	100.0±0.1	100.0±0.0	99.5±0.5
	1 mm ant	90.3±3.5	96.8±1.4	98.6±0.9	99.8±0.3	99.9±0.2	99.5±0.6	90.4±3.3	96.8±1.3	98.7±0.9	99.8±0.3	99.9±0.3	99.5±0.4
Introduced MLC error for coronal orientation/mm	*0*	87.5±3.8	93.9±3.3	98.0±2.1	99.1±1.1	99.5±0.8	98.1±0.8	87.4±4.3	94.1±3.5	97.8±2.1	99.0±1.0	99.6±0.7	98.1±1.1
	0.2	81.7±6.5	89.4±5.0	95.6±4.7	97.5±3.3	98.4±2.8	96.0±3.0	84.5±4.4	92.0±3.8	97.2±2.3	98.8±1.3	99.5±0.8	97.4±1.6
	0.4	78.0±4.7	88.3±2.8	95.8±2.4	98.1±1.3	99.2±0.9	96.3±0.8	80.0±5.8	89.2±3.7	96.2±2.6	98.3±1.4	99.3±1.0	96.8±1.1
	0.6	72.1±5.3	84.1±2.8	94.5±2.5	97.4±2.0	98.9±1.0	94.7±1.0	74.3±6.5	85.3±3.9	94.6±2.7	97.2±2.3	99.0±1.1	94.8±1.7
	0.8	65.4±5.8	79.3±2.9	92.2±2.9	96.5±2.1	98.4±1.2	93.1±1.1	68.0±6.9	80.6±4.6	92.7±3.6	96.6±2.7	98.6±1.5	93.6±1.9
	1	59.9±6.5	76.1±3.3	88.7±3.6	94.4±2.3	97.9±1.4	91.5±1.0	62.3±6.5	77.5±4.5	89.3±4.6	94.6±3.1	98.3±1.7	91.7±1.9
Introduced MLC error for sagittal orientation/mm	*0*	88.5±2.6	95.4±2.1	98.9±0.6	99.8±0.2	99.9±0.1	98.8±1.2	88.8±3.3	95.6±2.5	98.9±0.8	99.9±0.2	100.0±0.1	98.9±1.3
	0.2	86.0±4.0	93.8±3.0	98.7±0.6	99.6±0.3	99.8±0.2	98.0±1.8	86.4±3.8	93.9±2.9	98.9±0.7	99.7±0.4	99.9±0.2	98.1±1.8
	0.4	82.5±5.5	91.6±4.0	98.1±1.4	99.4±0.7	99.8±0.2	97.4±1.8	84.2±6.5	91.8±3.5	98.5±0.9	99.6±0.3	99.8±0.2	97.5±1.7
	0.6	78.6±6.3	89.0±4.6	97.0±2.0	99.0±1.0	99.7±0.3	96.4±2.0	78.7±5.8	89.5±4.1	97.4±1.2	99.2±0.5	99.8±0.2	96.5±2.1
	0.8	73.6±6.6	85.7±4.9	95.4±2.7	97.7±1.8	99.4±0.5	95.4±2.0	73.2±6.4	86.0±4.7	95.7±2.2	97.9±1.6	99.6±0.4	95.5±2.0
	1	68.6±7.1	82.3±5.1	93.7±3.3	96.9±2.0	99.2±0.8	94.2±2.3	68.7±7.1	82.1±5.2	93.8±2.8	97.1±1.7	99.4±0.8	94.2±2.4

		Gamma pass rate for merged results with no SNC shifts applied/%						**Mean central diode disparity/%**
		1%/1mm	*2%/1mm*	*2%/2mm*	*3%/2mm*	*3%/3mm*	*5%/1mm*	
Introduced isocentre shift for coronal orientation	*0*	86.7±3.2	93.6±2.9	98.2±1.7	99.2±0.9	99.6±0.8	97.7±1.4	0.1±0.8
	1 mm left	72.7±7.7	82.2±7.3	94.1±5.9	96.6±3.8	98.9±1.7	92.7±3.8	0.0±0.9
	1 mm right	74.9±9.0	85.5±6.3	95.8±3.5	97.7±2.7	99.1±0.9	94.3±4.0	0.1±0.8
	1 mm sup	86.4±3.7	92.9±2.5	97.6±1.5	98.8±0.9	99.4±0.9	97.9±0.8	0.0±1.0
	1 mm inf	61.8±7.6	73.5±6.9	85.5±5.0	90.1±4.1	96.8±3.4	87.9±4.1	0.0±0.9
	1 mm post	78.6±5.2	86.8±4.1	96.3±2.2	98.4±1.1	99.3±0.8	95.3±1.9	0.2±0.8
	1 mm ant	80.8±6.9	89.3±5.6	97.4±1.8	98.7±1.0	99.3±0.8	95.8±2.5	0.2±0.8
Introduced isocentre shift for sagittal orientation	*0*	81.0±9.2	89.9±7.4	97.0±2.3	98.6±1.8	99.9±0.2	96.7±3.4	-1.2±0.9
	1 mm left	85.5±5.9	94.0±3.0	98.8±0.8	99.7±0.3	99.9±0.2	98.5±1.4	-1.3±1.0
	1 mm right	84.5±5.3	92.7±2.9	98.2±0.9	99.6±0.3	99.9±0.1	97.8±1.4	-1.2±0.9
	1 mm sup	87.5±3.8	94.9±2.1	99.0±0.7	99.9±0.1	100.0±0.1	98.9±0.6	-1.3±1.0
	1 mm inf	72.7±7.5	85.8±6.5	97.1±1.7	98.8±1.0	99.8±0.2	94.4±2.7	-1.2±0.9
	1 mm post	85.3±2.2	93.3±1.4	98.6±0.6	99.7±0.1	99.9±0.1	98.3±0.8	-1.3±0.9
	1 mm ant	80.1±6.8	89.9±4.8	98.0±0.9	99.4±0.5	100.0±0.0	96.6±2.6	-1.4±1.1
Introduced MLC error for coronal orientation/mm	*0*							0.1±0.8
	0.2							0.2±0.8
	0.4							0.3±0.8
	0.6							0.3±0.8
	0.8							0.4±0.8
	1							0.5±0.8
Introduced MLC error for sagittal orientation/mm	*0*							-1.2±0.9
	0.2							-1.2±0.9
	0.4							-1.2±0.8
	0.6							-1.0±0.8
	0.8							-0.9±0.7
	1							-0.8±0.7

There are several key points that can be derived from this table. Firstly, in general, the introduction of isocentre shifts has little effect on the mean gamma pass rate for each set of criteria, when compared to the introduction of MLC errors. The result of this is seen in Section 3.5, where the sensitivity and specificity of MapCHECK to introduced isocentre errors is shown to be worse than random guessing for several gamma criteria. The effect of including automatic shifts in SNC Patient software is a general increase in the pass rate for all gamma criteria. This has repercussions which can be observed in the ROC curve results for these criteria (refer to Section 3.5).

In terms of gamma pass rate, the effect of doubling the effective resolution of MapCHECK by merging two measurements has no discernible effect on the final results. All mean gamma pass rates have overlapping error brackets of 1 SD for each combination of gamma criteria and error, suggesting that there is no significant increase in accuracy when detection resolution is increased. Based on these results, increasing the resolution is only necessary when one is looking for points measured in the high dose gradient region, as these points can be vital for determining why the plan failed QA. For non-stereotactic plans, increasing the resolution of MapCHECK will probably not provide any additional valuable information to the physicist, since the sample size provided by the standard MapCHECK diode resolution is probably sufficient for these plans.

All mean central diode disparities listed were below ± 1.5% from the expected result. The maximum disparity recorded for any individual central diode reading was -2.83%. Point dose measurements are expected to be within ± 3.0% of the expected measurement [83, 84], hence the results obtained here indicated acceptable agreement for commissioning standards that pertain to absolute dosimetry.

3.5 ROC curve analysis

Figures 3.14 and 3.15 show the ROC curves generated using the ROCR package, via comparison between the gamma analysis results and the actual DVH metric standards.

Based on Figure 3.14, it appears that MapCHECK in MapPHAN is proficient at detecting introduced MLC errors, with sensitivity and specifity comparable to that seen in the literature. The shapes of the curves in both coronal and sagittal orientations are relatively similar, for their respective gamma criteria. For instance, the results for 3%/3mm show a slight dip below the line of

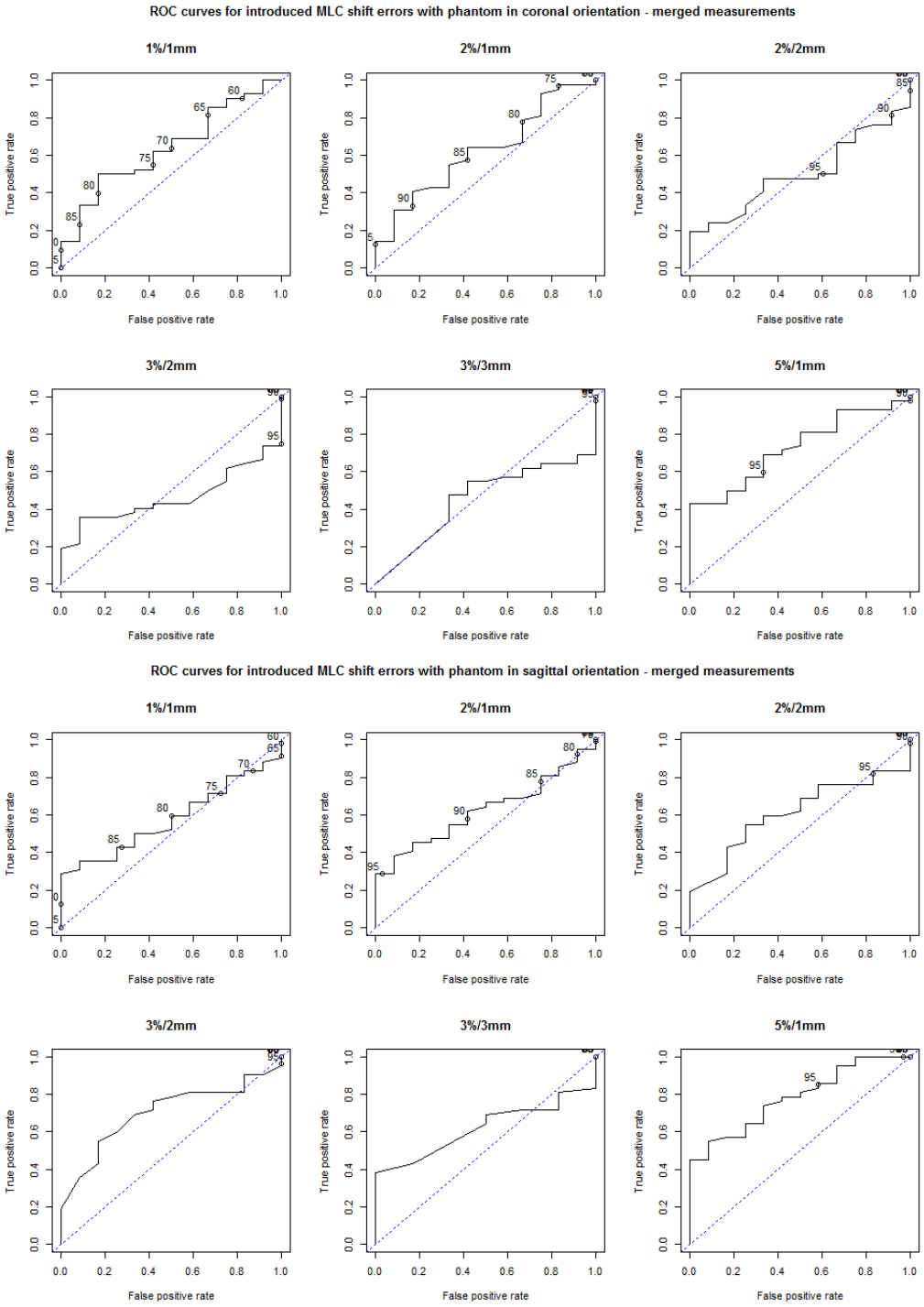


Figure 3.14: ROC curves generated from merged MapPHAN results with multiple gamma analysis criteria, for plans with introduced MLC shift errors delivered to MapPHAN in both coronal (top) and sagittal (bottom) orientations.

$y = x$ for both orientations, although the magnitude of this dip is greater for the coronal orientation results. Similar trends can be seen for the other criteria. Hence, assuming that variations in these trends are statistical in nature, it can be concluded that neither phantom orientation produces superior sensitivity or specificity for the detection of introduced MLC errors. Since MLC shifts are

performed in an axis orthogonal to both of these phantom orientations, this result makes sense; in any case, the rotational nature of delivery ensures that the dose variation is spread out relatively isotropically around all edges of the target volume, as seen in Figure 3.1.

The results in Figure 3.15, on the other hand, suggest that the phantom exhibits poor sensitivity and specificity to introduced isocentre shift errors. The ROC curves tend to fall below the line of $y = x$, suggesting that the phantom may be performing worse than random guessing for detection of these errors. These were unexpected results, but multiple secondary checks of individual classifier points were performed to validate the charts. Assuming that the tests are not applying information incorrectly, it can be tentatively concluded that these results are statistical aberrations. This is plausible, considering the low data density for the isocentre shift results. Due to the increased risk of linac collisions when performing QA, some patient plans were removed from the analysis. More plans were removed from QA for the sagittal case, since this phantom orientation increased the risk of collisions significantly. Hence any quantitative conclusions derived from these results carry less weight than those derived from the MLC shift results.

Additionally, the AUC for each curve was calculated in the ROCR package. These results are shown in Table 3.4.

Table 3.4: AUC for ROC curves for both isocentre and MLC shift errors, using all 6 gamma criteria. Both merged and unmerged results are shown, with the phantom in both coronal and sagittal orientations. The maximum discrepancy between any pair of merged and unmerged results in this table is 0.16, while the mean discrepancy between pairs in this table is 0.07 with a SD of 0.04.

	Error type	Phantom orientation	1%/1mm	2%/1mm	2%/2mm	3%/2mm	3%/3mm	5%/1mm
Merged	Isocentre	Coronal	0.15	0.18	0.15	0.22	0.18	0.42
	Isocentre	Sagittal	0.52	0.67	0.43	0.48	0.32	0.64
	MLC	Coronal	0.64	0.62	0.50	0.46	0.45	0.73
	MLC	Sagittal	0.58	0.62	0.61	0.70	0.62	0.78
Unmerged	Isocentre	Coronal	0.13	0.18	0.19	0.29	0.20	0.46
	Isocentre	Sagittal	0.66	0.59	0.28	0.32	0.44	0.74
	MLC	Coronal	0.54	0.50	0.43	0.38	0.38	0.61
	MLC	Sagittal	0.64	0.67	0.60	0.71	0.61	0.88

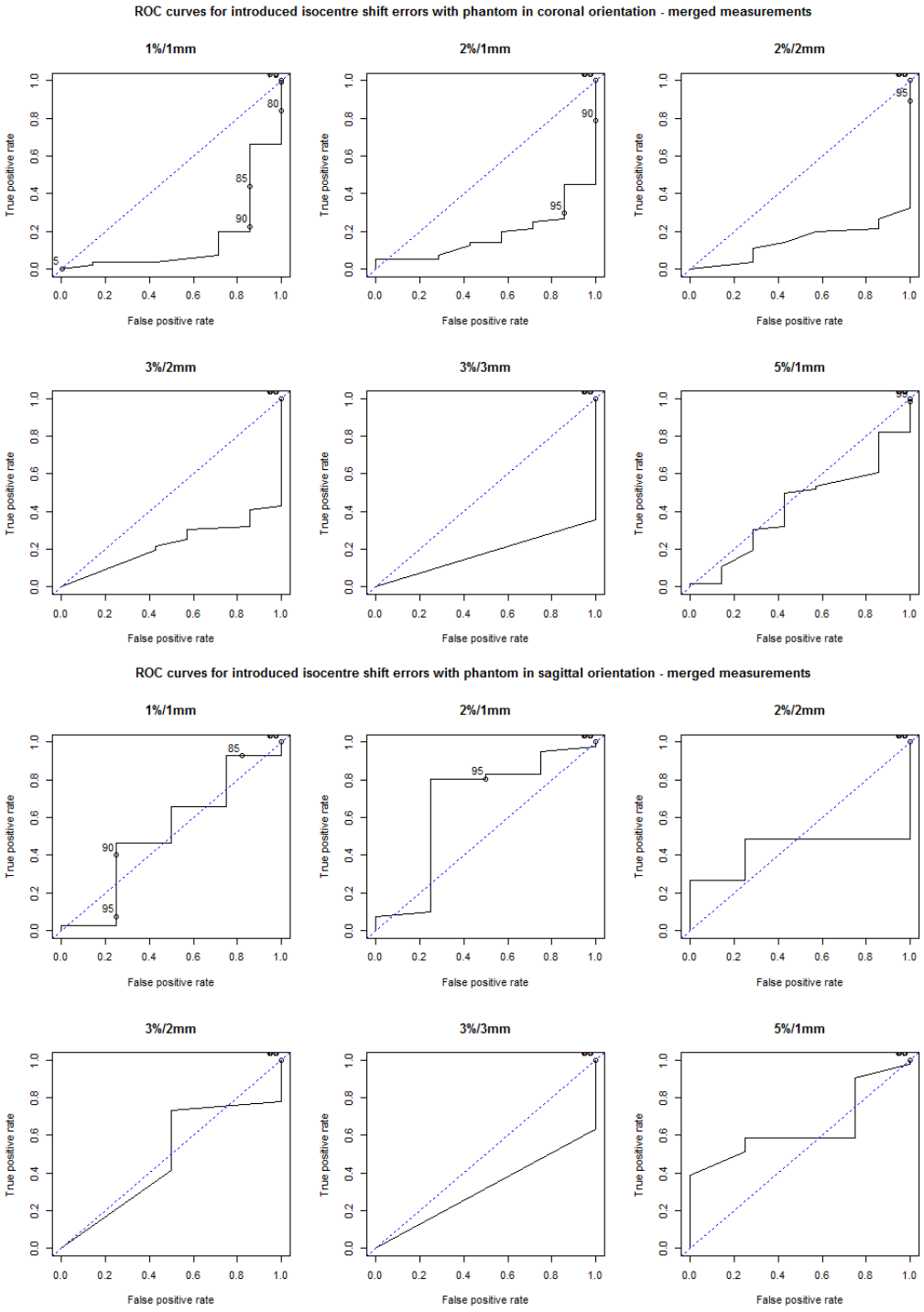


Figure 3.15: ROC curves generated from merged MapPHAN results with multiple gamma analysis criteria, for plans with introduced isocentre shift errors delivered to MapPHAN in both coronal (top) and sagittal (bottom) orientations.

3.5.1 Comparison of merged and unmerged results

The results shown in Figures 3.14 and 3.15 were generated using the merged measurements, in which the dose was measured twice - the second time with the phantom shifted 5 mm superior to the

original phantom position. ROC curves were also generated for the unmerged results, where measurements were only taken with the phantom in the original position on the couch. These can be seen in Figures 10 and 11 in the Appendix, while the AUC of these curves is included in Table 3.4. The low mean discrepancy between the pairs of merged and unmerged AUC results is reflected in the low mean discrepancy observed between the pairs of merged and unmerged gamma analysis results.

3.5.2 Application of automatic software corrections for isocentre shift results

Initially, SNC shifts were calculated for all plans, as per ADHB protocol. However, the ROC curves generated for the isocentre shift errors suggested that inclusion of these automated corrections was reducing both the sensitivity and specificity of the phantom to these geometrical errors. In order to determine if this was the case, the computed and measured plans were compared again in SNC Patient, this time without the application of shifts. The results of this analysis can be observed in Figure 3.16. Once again, the AUC for each curve was calculated, as shown in Table 3.4.

There is a significant improvement for the results obtained with the phantom in the coronal orientation, as seen in the mean increase in AUC of 0.19 for measurements taken without automatic shift corrections. However, in the sagittal orientation, the sensitivity and specificity actually decreases overall, with a mean decrease in AUC of 0.16 for measurements taken without automatic shift corrections. It is likely that this is due to the low data density in the regions of the curve that contribute most significantly to the AUC. In general, the 1 mm isocentre shifts did not generate significant dosimetric fluctuations, meaning that the majority of datapoints contributed to the point (1,1) in ROC space (representing perfect sensitivity but zero specificity). Data that contributes to this point in ROC space does not affect the calculation of AUC in any way. Considering that the AUC is only an indication of the sensitivity and specificity of gamma criteria as a whole, the specific thresholds which would be applied clinically are not being specifically analysed here. Assuming the data for the phantom in the coronal orientation is statistically significant, it should be concluded that including shifts in SNC Patient significantly reduces the sensitivity and specificity of the phantom to introduced geometrical errors.

Table 3.5: Comparison of AUC results for gamma checks performed with and without shifts in SNC.

Shifts applied	Phantom orientation	1%/1mm	2%/1mm	2%/2mm	3%/2mm	3%/3mm	5%/1mm
No	Coronal	0.42	0.44	0.35	0.41	0.29	0.57
	Sagittal	0.27	0.37	0.25	0.40	0.44	0.34
Yes	Coronal	0.15	0.18	0.15	0.22	0.18	0.42
	Sagittal	0.52	0.67	0.43	0.48	0.32	0.64

Correlation between gamma pass rates and AUC

Figure 3.17 shows the mean gamma pass rates for each type of introduced error, for all gamma thresholds used during this experiment. The error bars in this chart can be interpreted as the degree to which the introduced error changed the gamma pass rate. For instance, the variance in the 1%/1 mm results due to introduced MLC errors shows that the introduction of these errors generated a large change in the gamma pass rate; this is confirmed in Table 3.3. The figures also confirm that the introduction of automatic shifts in SNC significantly increases the mean gamma pass rate for isocentre shift errors.

An attempt to determine any potential correlation between mean gamma pass rate with the AUC for each level of introduced error was made. The results for this analysis were not statistically relevant due to the low number of datapoints available, but still provided some insight into the nature of potential relationships between the pass rate and AUC. This is discussed further in Section 4.5.1.

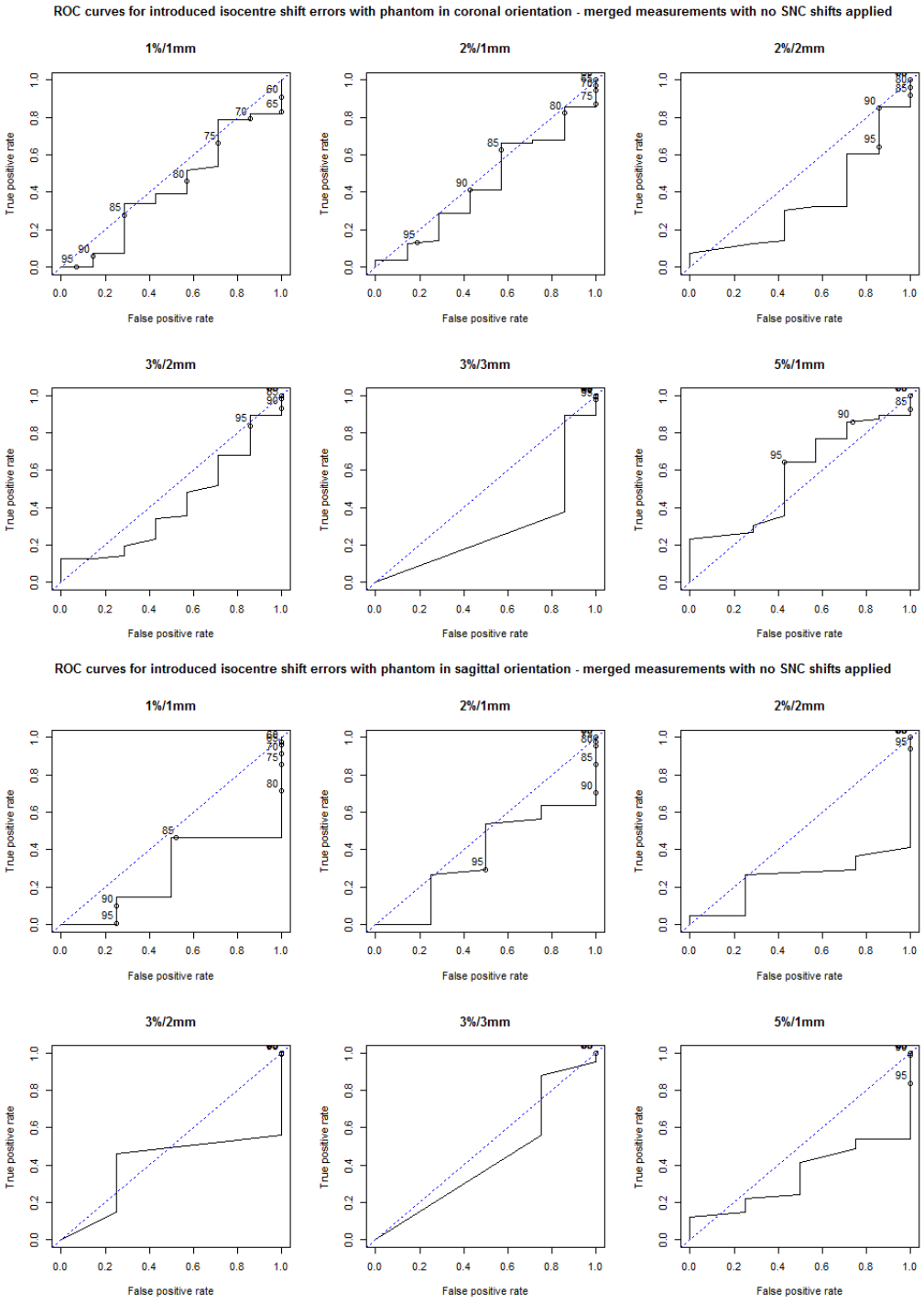


Figure 3.16: ROC curves generated from merged MapPHAN results, with no SNC shifts applied, for plans with introduced isocentre shift errors delivered to MapPHAN in both coronal (top) and sagittal (bottom) orientations.

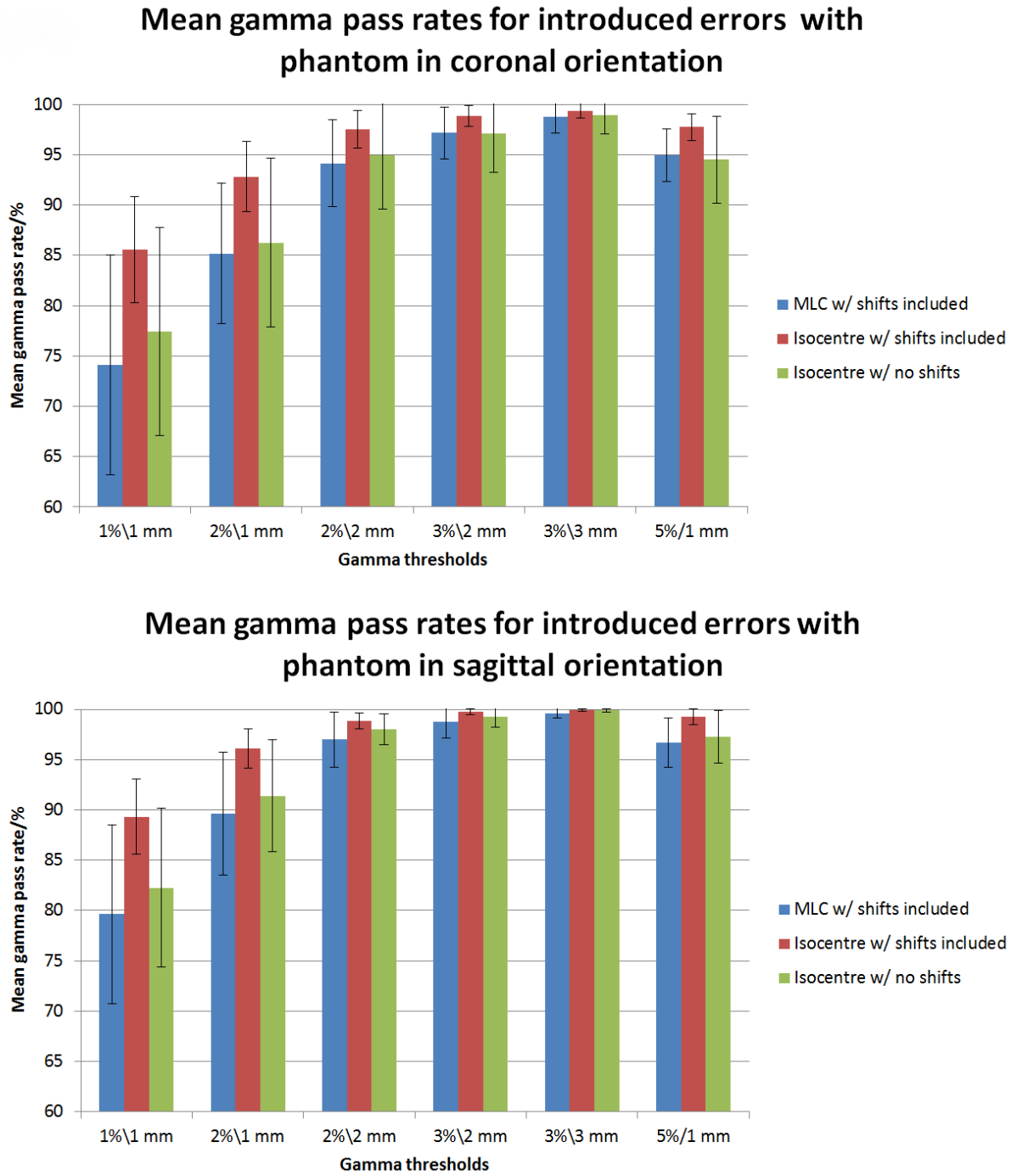


Figure 3.17: Mean gamma pass rates for all introduced errors (MLC and isocentre shift errors), with phantom in coronal and sagittal orientations.

Chapter 4

Discussion

Several studies analysing PTV doses and associated QA have been previously carried out [63–68]. This thesis focusses instead on the QA of OAR doses, a topic which is scarcely covered in the literature. As far as the author knows, the closest study of this type was performed by Kim *et al.*, who used ROC curves to analyse the sensitivity of the gamma index to introduced errors, but did not focus on OAR doses, and only studied MLC shift errors [65]. This thesis reinforces some of the conclusions of this paper, while introducing several unique concepts.

4.1 *In silico* analysis of introduced errors

The choice of machine errors to introduce into the sensitivity analysis was mainly determined by the predicted influence these errors would have on the treatment plan and resultant dose distribution. A literature review suggested that introducing errors in MLC position and performing isocentre shifts would have the most significant influence on plan outcomes [63–68, 78, 85, 86]. In addition, the magnitude of the errors was limited to the machine QA tolerances applied at Auckland Hospital, which made the results both efficient in their determination and practical for application at this centre. For the MLC errors, the magnitude of the resultant dosimetric changes was significant within these tolerances, perhaps suggesting that a review of these tolerances could be appropriate. However, studies focussing on PTV dose have indicated that gamma analysis is not sensitive to isocentre shift errors less than 2 mm [63]. If anything, this study on OAR doses does not disprove this outcome.

It is worth noting that in reality, every treatment plan has some level of error in every continuously definable parameter. Therefore, statistically, some plans will be afflicted by a combination of errors that results in significant dosimetric disparities, while the combination of errors for other plans will result in smaller disparities. This is why QA is an absolute necessity, in order to identify the plans that are most afflicted by combined errors. MLC position and isocentre shift errors comprise only two of these parameters, but represent a large proportion of the entire scope of potentially deliverable treatment errors. Hence as an analytical tool, the introduction of these types of error can provide a good indication of the robustness of the quality assurance process, especially when the errors are pushed to the extent of their tolerances.

MLC leaf position errors have been studied on more complex IMRT plans, but to the author's knowledge, this is the first study that looks at conformal arc plans for lung SABR specifically, with a focus on OAR DVH metrics, rather than on the PTV. The lack of OAR analysis in the literature is an indication of the priority placed on PTV coverage during QA. Granted, the PTV always takes highest priority in a radiotherapy plan. However, the priority placed on particular OAR dose constraints is such that they will shape the treatment plan to a significant degree. It seems logical that PSQA processes should place a similar (if not identical) priority level on the QA of OAR doses. Ideally the physicist is also performing a visual inspection of the shape of the dose distribution, although it can be all too easy to accept a plan that passes all gamma analysis criteria by themselves. Certain QA techniques provide this level of analysis, showing hot and cold spots in the plan on the original CT dataset, and allowing the physicist to perform a more thorough analysis of doses to different volumes of tissue. SNC Patient software does not achieve this level of detail; ideally it should be used in conjunction with another QA technique in order to achieve this.

It should be noted that doctors will regularly place priority on PTV coverage at the expense of delivering excessive dose to certain OARs. For instance, it is common clinical practice to permit excessive dose to be delivered to the ribs and chest wall during lung SABR, significantly increasing the risk of rib fractures. This is done to prevent excessive constraints from being placed on the treatment plan, which then allows the planner to improve PTV coverage and potentially reduce dose to more critical OARs such as the heart and spine. Hence while it should be concerning that these OARS are receiving excessive dose, there is certainly justification for these doses in many cases.

In retrospect, the choice of conformal arcs for measurement was useful for determining the utility of

the phantom for rotational deliveries. The simplicity of conformal arc deliveries allowed the detection of certain errors with greater confidence as to their origin, which would have been more difficult with the smaller MLC field sizes and intensity modulation of VMAT. An example of such an error was covered in Section 3.2.1. The detection of the offset between calculated and measured dose at the proximal end of the phantom (relative to the mean positional origin of radiation delivery) could clearly be attributed to the phantom structure, as it was appreciated that the dose distribution was weighted towards the end of the phantom closest to the mean origin of radiation delivery. The addition of intensity modulation would likely have shifted the weighting of radiation delivery, making it more difficult to ascertain the source of error.

Estimation of standard dosimetric error

Mzenda *et al.* performed a comprehensive analysis of the dosimetric equivalence of Pinnacle and Raystation calculations, as part of the commissioning process for Raystation at Auckland Radiation Oncology, in a study that was performed with help from members at Auckland Hospital [87]. They found that open field calculations performed with Raystation agreed with measured results to within 2% for central axis points, within 10% for measurements taken in penumbral regions, and within 4% for off-axis measurements, quoted to within ± 2 SD. Any point of maximum dose in an OAR adjacent to the PTV will, at various stages of the treatment, find itself in all three of the beam regions listed above. Hence the dosimetric uncertainty at this point should be calculated as an appropriated ratio of the general uncertainties in these regions. Several approximations in the beam model, and the potential disparity between the accuracy of the beam model at ADHB compared to that used in Mzenda's paper, will modify the uncertainties listed above. Therefore it is appropriate to estimate a conservative standard uncertainty for all results obtained in the following charts. Hence an 8% standard dosimetric uncertainty was considered appropriate as an estimated uncertainty in all gradients plotted, quoted to within ± 2 SD. This also assumes a higher uncertainty in calculating dose in low density lung tissue, due to poor modelling of electron transport using the collapsed cone convolution algorithm in Raystation [87].

4.1.1 DVH measurements

The trends obtained in Figures 3.5 through 3.8 were not unexpected. The initial maximum point dose received by an OAR is, to a certain degree, representative of the proximity of the OAR to the target volume. This becomes apparent when one looks at Figures 3.1 and 3.2. The most significant dosimetric increases are observed around the periphery of the PTV, meaning that OARs with tissue in these regions were most significantly affected by introduced errors.

The choice of DVH metrics for measurement were based on a template for treatment plans at Auckland Hospital, as this allowed the direct analysis of the QA processes at this centre. By comparing the change in maximum point dose to the initial maximum point dose received by an OAR, it was possible to eliminate any uncertainties surrounding geometrical metrics. Initially, measuring the change in maximum point dose as a function of distance between the PTV and the specified OAR was considered, but this would have been problematic, since the choice of points used to measure distance was critical. As this was an OAR based study, points had to be chosen that best reflected the dose being received by individual OARs. The centroid point of the contoured organ could have been used, but in many cases this would have been an unfair descriptor of the organ's position, due to the variable nature of organ shapes. As such, the centroid position is not a fair reflection of the dose being received at all points in the organ, most notably since its location is determined by the planner's contouring of the organ.

Alternatively, the dose received by the point of the organ closest to the PTV could have been chosen, but this also was problematic, as the point of location of the PTV itself is often offset from the location of maximum dose inside the PTV, especially when the PTV is non-spherical in shape. Even if the point of maximum dose was used as a centre to establish the distance of OARs from the PTV, the shape of the dose distribution itself would have been a factor introducing variation in the trend, since the dose gradient is typically sharper in the crossplane direction for conformal arc treatments. Hence the change in dose was measured as a function of the maximum point dose received by the organ, which provided a semi-reliable indication of the effect of introduced errors.

However, using this metric brings its own series of uncertainties, the first being the gradient of the trend itself. As mentioned previously, the introduction of a 0.6 mm MLC error generates an unexpected dip in the calculated maximum point doses for almost all OARs. Since the source of this

error is probably associated with the dose grid, it is difficult to eliminate entirely, although it would be unfair to exclude it from the final results. Secondly, and more importantly, the change in maximum point dose remains a function of the spatial geometry of the organ relative to itself and to the PTV. The maximum point dose received by an organ must be a function of the organ's position in the dose distribution, but the point of maximum dose may change as errors are introduced into the treatment plan. This is obviously the case for isocentre shifts, where the shift in dose is in the direction of the introduced isocentre error. If the phantom geometry is ignored entirely, then the expected change in dose as a result of shifting the isocentre can be approximated as the differential of the dose at that point, as a function of the shift. Critically, this means that high dose gradient regions have the greatest uncertainty when comparing measured and calculated dose distributions, using gamma analysis or similar. Hence the dose to OARs located in these regions should be more carefully managed, especially if they are close to recommended threshold doses.

A similar trend is observed in MLC results, although a mathematical basis for the trend cannot be formulated as simply as for isocentre shifts. The most important similarity between the two trends is the increase in uncertainty for regions with high dose gradients. In the case of MLC shifts, these regions lie approximately underneath the edge of the MLC leaves (using a beam's eye view, as in Figure 2.1). Hence it would not be accurate to utilise the differentiated dose distribution to analyse the expected change in dose upon the introduction of an MLC shift error. However, this is an appropriate technique for determining where the greatest uncertainties in the dose distribution are located, since as with isocentre shifts, the greatest change will be in the high dose gradient regions. In order to develop an accurate model of the uncertainty in MLC errors, one would likely need to perform a meta-analysis of MLC errors in multiple linacs at different clinical centres. This would be an extremely time-intensive project that would probably not be worth the quantification it provides, as developments in MLC technology are continually reducing the magnitude of MLC errors.

Interpretation of maximum point dose

The results observed in this thesis highlight how impractical certain metrics can be for planning purposes. One of the main disadvantages of the DVH is the lack of spatial information that it provides, in that it does not let the user know where the dose is located. The use of maximum point dose seems like a particularly arbitrary threshold for measurement, as it provides no spatial

information about the OAR in question. The main utility of this metric is that it can be used for protection of serial organs, as the theoretical consequence of exceeding the tolerance dose at one point in the organ is failure of the organ as a whole. A prime example of this is the spinal column, which is comprised of millions of neuronal tracts, some of which extend from the brain stem all the way to the sacrum. Any damage inflicted to these tracts at any point in the cord will result in failure of the tract as a whole. However, when a maximum point dose exceeds tolerance in a plan, one should not simply accept this result at face value without considering the uncertainties involved. Where is the OAR in relation to the PTV? At what location in the OAR is the point of failure? What is the dose gradient at this point of failure? How far does the high dose region extend into the OAR? To what extent can the OAR be considered 'serial' as opposed to 'parallel', where failure at one point in the organ does not necessarily constitute failure of the organ as a whole? While the last question is more medically oriented, the prior questions all require an understanding of the dosimetric uncertainties involved in the planning process. This all goes to highlight the importance of not using the DVH as a standalone tool for plan assessment.

Additionally, in several cases, the CTV may be directly adjacent to OARs which then fall in the bounds of the PTV itself, following expansion of the CTV margins. In these cases, the OAR's maximum point dose tolerance will almost certainly be exceeded, making the metric virtually useless in providing information to the clinician. In these cases, it is vital that a metric is used that determines dose to the OAR outside of the PTV. One way of doing this is to contour the OAR in such a way as to exclude any section of the OAR that falls in the PTV. Even in these cases, however, the maximum point dose will certainly fall on the PTV-OAR boundary, which again renders the metric somewhat useless as a means of providing information about the OAR to the clinician. This highlights the fact that the isodose curves on the plan itself should be checked thoroughly, in addition to the use of relevant DVH metrics. The clinician should also have sufficient knowledge of the uncertainties involved in these plans, so that they can make a judgement call on dose metrics that have only just fallen within acceptable tolerances.

The complexity of the human body is such that most radiotherapy plans will fail at least one threshold for planning tolerance. Hence the planning process always involves some form of cost optimization process, whereby multiple sub-optimal solutions must be compared and contrasted in order to generate what the planner believes to be the most suitable plan. Inverse planning invokes a

multicriteria cost optimization technique known as Pareto optimization, whereby a human operator sets the boundary constraints on a computational optimizer, which determines the Pareto optimal point on the boundary of the objective space that the operator has set. If the objective functions are optimised correctly, then this point will allow the planner to achieve the highest possible plan quality for those limits, within the bounds of reasonable time constraints [42]. It can be safely assumed that almost all complex plans will push (and sometimes exceed) their specified OAR tolerances in order to maximise PTV coverage. It is then left to the clinician to determine which tolerances are acceptable to push. This is where a complete understanding of the uncertainties involved in these metrics is invaluable.

Dose grid resolution

One parameter that may have some effect on the maximum point dose metric is dose grid resolution. For any single OAR, Raystation will collate a matrix of dose voxels within the dose grid, and calculate the average dose in each voxel. Since the maximum point dose necessarily increases as the size of each voxel decreases, it is possible that the resolution will have an effect on this metric. Park *et al.* examined the optimal dose grid size to use for practical dose calculations using DCAT in lung SABR. When compared to a 2 mm dose grid size, they found that using a 4 mm dose grid size increased the dose variation by up to 3-4% (approximately 50 cGy) for OARs, while a 3 mm dose grid size produced a variation of less than 1% (approximately 12 cGy) [88]. This suggests that below a dose grid size of approximately 3 mm, the dosimetric dependence on dose grid size is negligible.

However, the position of the dose grid also appears to have some effect on the results. Chung *et al.* performed a phantom study, in which they prescribed 5,400 cGy to shallow and deep targets, using Pinnacle. They found that dose grid size had a significant effect on the prescribed dose to the PTV [89]. When this error is combined with the variation introduced by changing the dose grid resolution, the choice of dose grid parameters becomes significant. Even in this thesis, some dose grid effects were observed, as seen in Section 3.1.2. While these effects may have been relatively small compared to the overall trend, it still highlights the importance of consistency in dose grid size, as a larger dose grid may have generated a slightly different trend.

In summary, the uncertainties present in calculating the maximum point dose for a volume are significant. When this metric is then used for plan evaluation, it has the potential to be misused. If

the planner or clinician does not appreciate the uncertainties inherent to the metric, they will be more prone to misinterpretation of a borderline result.

4.1.2 Utility in treatment planning

Several modern TPS's, including Raystation and Pinnacle, will place a 'warning' flag on evaluation metrics that are close to failure. The recommended course of action one should take upon seeing these flags is often somewhat unclear, and is left to the discretion of the clinician, increasing the potential for human error in clinical plans. Those with a proper appreciation of uncertainties may view this as a recommendation to place extra restrictions on the dose to the affected volume. Others may simply read the metric as a warning, and accept the result without considering the probability of the actual patient dose exceeding the calculated value.

The author would like to propose an alternative solution: a confidence interval that could be introduced to the maximum point dose for any volume. The interval type should be relatively common - for example, one consisting of two standard deviations from the mean (95% confidence). The size of the interval would primarily be a function of the dose registered, as well as the position of the OAR relative to the treatment volume. If any OAR doses fell within one of these confidence intervals for a particular DVH threshold, then a recommended action could be taken by the planner, such as a preferential reduction of dose to that OAR. Of course, this would be a difficult solution to implement. In order to quantify the size of these intervals, a robustness analysis would need to be performed on a broader range of treatment errors than the ones covered in this thesis. The interval size would also be accompanied by its own error, which would need to be negligible compared to the interval size for this technique to be viable. The data density required for this degree of accurate prediction of interval size would therefore be sizeable. In addition, the interval size would certainly be a function of several other variables, such as linac type, treatment site, and beam energy, amongst others. It would also be necessary to perform the robustness analysis using multiple dose grid resolutions, in order to quantify their effect on the interval. In conclusion, this technique may not be viable for implementation in an entire TPS, much less a standard clinical centre. However, this shows just how much information is missing when one interprets the maximum point dose as an evaluation metric.

An alternative to the above solution would be to quantify the predicted change in NTCP values for introduced errors. This would provide a metric that is more clearly understood by a broader audience, as the NTCP for any particular OAR is simply a percentage value. Unfortunately, the uncertainties inherent to NTCP calculations are even more significant than dosimetric uncertainties due to the lack of clinical data available. (Additionally, the magnitude of the TCP uncertainties must be, by definition, a function of the magnitude of the uncertainties in dosimetry.) Still, the end result would be more readily interpreted if it were a simple NTCP percentage rather than an uncertainty bracket.

A simpler alternative to using maximum point dose is a scaleable volume metric, such as $V_{30\ Gy}$ which was used for the chest wall in these experiments. The DVH already provides this functionality inherently. This type of metric is determined from a number of voxels that is high enough to negate the effect of both dose grid resolution and position. For instance, while an average dose of the entire volume can be measured with this technique, the average dose of far smaller volumes can also be measured, while still covering a large number of voxels. A compromise between using a point dose metric and a larger volume-based metric can thus be achieved.

It should be noted that during the writing of this thesis, Auckland Hospital updated its DVH metrics for conformal arc lung SABR plans to include both 'mandatory' and 'optimal' constraints, based on international recommendations in trials such as CHISEL, RTOG 0915, and SAFRON II [90–92]. This is at least a partial reflection of the uncertainties involved in treatment delivery, although the decision of which constraint to use may still be open to interpretation.

4.2 Commissioning of MapPHAN

The angular dependence of MapCHECK inside MapPHAN was successfully reduced to levels similar to those achieved by Jursinic *et al.* [75]. Since the MapCHECK model used in this thesis is different to the ones used in their paper, the results aren't directly comparable. However, the similarities between the angular dependency trends are sufficient indication that the techniques used here to minimise angular dependency could be applied to a broad variety of phantoms with planar (and perhaps even non-planar) detector arrays.

4.2.1 Angular dependency of diodes and determination of density overrides

Based on the results of this thesis, the recommendations that Sun Nuclear provide for commissioning of MapPHAN appear inadequate for accurate QA of arc therapy. This conclusion has been reached in other papers [75, 76, 93, 94], although suggested alternatives to correct for this issue are scarce and variable. The methods used in this thesis are most closely aligned with those utilised by Jursinic *et al.* [75], based on the literature sampled.

A 10^{o} gantry spacing was considered acceptable for these measurements, assuming that any potential aliasing effects would be minimal, as this spacing is on par with measurements taken in the literature. In addition, Jursinic *et al.* [75] took measurements using only the central diode, and attained an angular dependency that was within 2%. The results in this thesis were taken using the mean dose calculated to the central five diodes, making the assumption that the dose falloff over this 1 x 1 cm^2 region would be approximately linear. This technique increased the statistical relevance of the results, and was an appropriate surrogate for the central diode dose. However, it also provides a more realistic assessment of the angular dependency of the phantom structure as a whole, rather than analysing the dependency of the central diode alone, which has been performed multiple times before. By analysing more than one diode simultaneously, the effect of systematic biases in the dependency of the central diode were reduced.

The measurements taken in the central diode region (within 40 mm of the central diode column) should be given priority in commissioning this device for lung SABR, for two reasons: the first being the fact that 10 x 10 cm^2 beams were used for the angular dependency measurements, meaning that diodes at a distance $\geq$ 50 mm from the central diode were either in the buildup region or outside of the central beam axis region for a high proportion of the measurements. The uncertainty in these regions has been found to be significantly higher than for measurements in the central beam axis region [87], which is reflected in the data here. Secondly, since the device was being commissioned for the QA of lung SABR plans, it can be assumed that the maximum PTV size is below 8 cm in width along its major axis, meaning that the high dose-gradient SABR-generated dose distributions will fit within this central diode region. Because the SNC Patient software was configured to ignore doses below a threshold of 10%, the only measurements that are truly relevant to the QA of lung

SABR plans are those in close proximity to this central diode region. The caveat of this is that the device will need to be re-commissioned before it can be used for the QA of treatment plans with fields that extend beyond this central diode region.

By changing the density of the array region and measuring the ratio of calculated to measured dose for other gantry angles, it became clear that the density of the array only had a significant effect when the beam was passing through it at gantry angles from 80° to 100°, or from 260° to 280°, where the effective thickness of the array relative to the beam axis was greatest. This meant that the array density override could be calibrated relatively independently, using just these gantry angle measurements to determine the most appropriate density to use.

It was decided to set the array density in such a way as to maximise alignment between the PDDs in the central region of the array, while sacrificing alignment towards the edges, as can be seen in Figure 3.11. This was permissible for multiple reasons. Firstly, when creating the QA plans, the centre of the array was always aligned with the region of maximum dose - i.e. the region of the target volume. For lung SABR, this volume is typically quite small, on the order of up to 5 cm. Hence the region of the array requiring the highest degree of calibration accuracy was the central 5 x 5 cm^2 portion which was going to be irradiated directly for each patient plan. The alignment of the PDDs over this region was within 1% tolerance for G90 and G270 beams, indicating that there would be no dosimetric disadvantage to aligning the PDDs in this way, when measuring in this central array region. Secondly, the range of gantry angles where the disparity is highest remains small compared to the entire range of angles used during a single treatment. The only gantry angles that were significantly influenced by the disparity were G90 and G270.

To illustrate this, changing the array density in Raystation from 0.75 g cm^{-3} to 0.80 g cm^{-3} led to a 2.2% change in calculated dose to the central diodes when the beam was delivered at gantry 90°, but only a 0.2% change in calculated dose at gantry 80° and 100°, and negligible difference for all other gantry angles. Hence even if the density of the central section were assigned incorrectly, the effect this would have had on the total calculated dose would have been relatively low. Considering the fraction of the arc subtended by these affected angles, it can be assumed that the dosimetric errors that would have been introduced as a result of this intervention will be relatively minor. There would also have been minimal effect on the periphery of the array, where the dose received is of little concern due to the limits placed on PTV size.

Standard deviation of dosimetric discrepancy as function of gantry angle

The standard deviation in dosimetric discrepancy for central axis diodes (see Figure 3.12) can be interpreted as another measure of the agreement between the calculated and measured dose. However, when compared to the mean discrepancy, this value provides a better indication of regions where the density overrides applied are unfairly weighted towards increasing the dosimetric accuracy for a particular subset of diodes. An example of this can be seen in Figure 3.11. The SD increases significantly from 1.35% to 6.02% when all central axis diodes, rather than just the central nine diodes, are included in the calculation. Most diode positions in Figure 3.12 show a similar - albeit less significant - increase in SD, with some exceptions (for instance, at G200 and G210 for the coronal MapPHAN). This highlights the fact that the density overrides applied during the angular dependency measurements were chosen to validate measurements in the central diode region, and may introduce dosimetric offsets to diodes closer to the phantom periphery. While some of the increase in uncertainty can be attributed to out-of-field doses, this is clearly not a factor for beams that pass parallel (or close to parallel) to the diode array. It is at these angles that the dosimetric utility of the phantom is weakest. This should be kept in mind by the physicist during QA. If they are delivering arc beams that are oriented primarily through the most affected range of gantry angles, then they can expect to see greater dosimetric uncertainty in their final dose distribution. The density overrides can't account for the potential of each diode having its own individual angular dependency curve, due to manufacturing differences between each diode. These differences are probably a factor in the significant variation seen in some of the outlier diodes observed in Figure 3.12. However, it is difficult to determine to what extent these outliers were also a function of the inaccurate application of density overrides for diodes at the periphery of the array. In essence, having these diodes housed in a phantom makes it virtually impossible to isolate the individual angular dependency characteristics of each diode. One would need to measure each diode's dependency in an external, isolated environment to observe the true variance of this effect.

Analysis of angular dependency in literature

The angular dependency trend observed in Figure 3.10 has been observed in several publications, who report a maximum change in sensitivity of up to 25% as a function of the angle of irradiation [54, 74–76]. Li *et al.* performed an angular dependency analysis on the Sunpoint diode

detectors used in ArcCHECK (identical in structure to those in MapCHECK 2) and plotted a graph of the discrepancy between calculated and measured dose for a number of gantry angles and diodes. The trends they found were similar in nature to the trends found here. In particular, the dip in measured dose for beams delivered at G90 and G270, with the associated increase in measured dose at gantry angles close to these cardinal angles, were both consistently observed. ArcCHECK houses 1386 diodes in a helical array, meaning that beams oriented at 90° and 270° to individual diodes would not have been directly impeded by other diodes before reaching the diode which was measuring the angular dependency curve. This confirms that the angular dependency observed is mainly due to the structure of the diodes themselves, and not the fact that the diodes are arranged in a planar array in MapCHECK 2. However, this almost certainly had some effect, as shown by the more pronounced fluctuations in the regions of Figure 3.10 where the beam passes laterally through the array.

Jursinic *et al.* state in their paper on angular dependencies in diodes that "Low energy electrons are backscattered at the interface of low and high atomic number materials, and this results in higher diode sensitivity to photons that enter from the directions of the front and back surfaces of the copper plate" [54]. In addition, the 'sandwich'-type construction of the array will likely produce heightened diode response when a lower fraction of the cross-sectional beam is directly attenuated by the array itself. Keeling *et al.* [74] found that this effect increased to a maximum potential discrepancy of 37% for beams passing through the diode array at 90° and 270°. Hence the inherent angular dependency of the diodes is likely compounded by the planar nature of the diode array, which affects beams with a central axis that is parallel to the plane of the array.

It should also be noted that this effect is more pronounced for lower energy photons, and has a greater dependence on field size for lower energy photons. This phenomenon was observed by Keeling *et al.* [74], who compared the angular dependence of MapCHECK's diode detectors in 6 MV and 10 MV photon beams, for field sizes of 5 x 5 cm^2 and 10 x 10 cm^2. A smaller field size was found to significantly increase the percent dose difference between measured and planned doses, with the increase in discrepancy being approximately proportional to the initial dose discrepancy when field size was decreased. The increase in discrepancy was also higher for 6 MV beams compared to that for 10 MV beams. This can be explained by lower energy beams having higher backscatter rates in media, resulting in an overresponse of the diodes to lower energy photons. Jursinic *et al.*

discovered that removal of the copper plate found in these Suncheck diodes produced as much as a 6.4% decrease in the angular dependencies of the diodes, and suggested that they should not be used for arc therapy QA [54].

In reality, stating the maximum angular dependence of the phantom as a single percentage is not a fair assessment of the entire angular dependency profile. While the mean discrepancy for the central diodes across all gantry angles is 0.28% for the coronal orientation and 0.39% for the sagittal orientation, there are certain gantry angles for which the standard deviation or mean discrepancy between measured and calculated dose (and sometimes both) is more significant. For instance, it can be expected that beams delivered between 180^o and 350^o with the phantom in the sagittal orientation can be expected to generate a mean discrepancy of 0.61% between calculated and measured doses for the central diodes. The statistics appear worse when considering individual diodes. For the same phantom orientation and beam, the mean discrepancy between calculated and measured dose for the diode at -10 mm from the central diode would be 1.93%. While this remains within the expected absolute dose tolerance of $\pm$ 3%, it is a concerning result, as these introduced dosimetric biases may not be fully appreciated by the physicist performing QA on a plan with beams predominantly in these angles. The assumption is made that the majority of plans are delivered with beams that cover the entire range of gantry angles, and that the dose being delivered is spread out evenly over the range of beam angles. In this way, the dosimetric imbalance generated at certain gantry angles is rectified somewhat by opposing imbalances at other gantry angles. This is a generalised solution that will not work for all plans, however.

In summary, the choice to optimize density overrides in Raystation to counter the angular dependencies in MapCHECK was suitable for commissioning purposes. The aim was to assess the angular dependency of the phantom in a more holistic manner than is seen in some of the literature. Stating the angular dependency from the central diode alone doesn't account for the fact that the corrections made to reduce the dependency were specific to this region of the phantom. If the other diodes were taken into account also, this would likely generate entirely different correctional values. By taking into account the central five diodes during angular dependency measurements, as well as looking at the dose to all diodes using SNC Patient software, both the inherent angular dependency of the diodes and the shape of the phantom as a whole could be accounted for.

SNC Patient software includes the angular dependence uncertainty in their measurements for

ArcCHECK diode measurements only, despite the fact that they designed the MapPHAN to be used for measurement of arc therapy. Even so, their stated value for this uncertainty is 0.5% after angular correction. This may be correct for ArcCHECK, but certainly not for MapCHECK. Even with the assumption that beams will be delivered to the phantom in directions orthogonal to the diode plane, the uncertainty still ranges up to ± 1% for beams within the ranges of G320-G40, and steadily increases beyond that (see Figure 3.9).

4.3 Patient specific quality assurance measurements

4.3.1 MapCHECK dose calibration

The initial results acquired for PSQA exhibited poor absolute dosimetry. MapCHECK is calibrated by delivering 100 MU of the desired beam energy to the centre of the phantom, with the diode plane located at isocentre in a coronal orientation. The dose recorded by the central diode of the array is then used as the calibration factor for calculating absolute dose in the remaining diodes. Initially, calibration was performed with a 3 cm slab of solid water placed on top of the MapCHECK phantom, as is convention at Auckland Hospital. Following calibration, MapCHECK was then inserted into MapPHAN and positioned for measurements. This contributed a mean dosimetric error of 2.0%, which had a significant effect on gamma pass rates.

A number of reasons for this discrepancy were posited, and various alterations were tested. A couple of plans were delivered with a reduced control point spacing (i.e. a reduction from 4° to 3° per control point), which had no significant effect on the dosimetry of the central diodes of the phantom. This result was expected, as control point spacing has the greatest dosimetric effect on the periphery, as can be seen in Figures 3.1 and 3.2, where the outline of the 40% isodose line can be observed. The jagged nature of this line is a direct result of the dose from individual beamlets being calculated at each control point. The isodose lines closer to the centre of the PTV are smoother due to the merging of these spikes generated by individual control point beamlets.

Two alternative beam models were then used to calculate dose in Raystation. One of these was an older version of the beam model which was not optimized for SABR delivery with smaller fields, and thus did not improve results at all. A newer beam model with a unique method of modelling

collimator scatter from the linac jaws and other beam modifiers was also tested. This beam model slightly improved the absolute dosimetry of the most central diodes, but was found to reduce the alignment between measured and calculated doses for the high gradient dose regions of the plan. Ultimately it was decided to use the original beam model for this reason.

It was discovered that the experimental setup used for dose calibration had a significant effect on the calibration value. An alternative dose calibration method was tested, with the phantom being inserted into the MapPHAN prior to measurement, providing 5 cm of solid water buildup instead of the usual 3 cm. Additionally, it provided additional backscatter due to the extra solid water placed underneath MapCHECK. This setup generated a dose calibration factor that was, on average, 2.0% higher than the factor generated by the use of 3 cm of solid water buildup. In order to correct the measurements previously taken with the incorrect dose calibration method, this 2.0% correction factor was retrospectively applied to the absolute recorded dose for the previously measured plans. The reasons for this discrepancy remain somewhat unclear, as the modelling of the variable levels of solid water in Raystation should, in theory, result in approximately identical dose calibration factors for both setups.

The issue is conflicted by the fact that the solid water used for the MapPHAN was modelled at 1.02 g cm^{-3}, which was shown to reduce the Raystation calculated dose to the central diode by 0.3% when compared to calculations performed with a density override of 1.00 g cm^{-3}. Because the density of the MapPHAN was determined based on the angular dependency measurements, there is some uncertainty in the accuracy of the density override of 1.02 g cm^{-3}. This value is likely to be offset from the true value in order to partially correct for the inherent angular dependencies of the diodes.

It is also possible that the presence of the solid water backing provides a significant addition of backscatter that is detected by the central diode but not accounted for by Raystation, as the collapsed cone algorithm in Raystation has been shown to model backscatter poorly [87]. The presence of a hole in the posterior plastic backing of the phantom, located directly beneath the central diode, would in theory increase the backscatter dose deposited in the central diode from the posterior piece of solid water. The rate of backscatter from this solid water is likely to be higher than that produced by the couch, which is lower in density with a decreased attenuation cross-section that, in theory, will produce a lower rate of backscatter as a result. Hence the presence of this piece of solid water is likely to have had a small effect on the dose calibration. In summary,

this highlights the importance of using the same setup conditions for dose calibration and measurement, if possible, as this prevents the possible addition of unaccounted factors.

Despite its issues, MapCHECK housed in MapPHAN has two key advantages as a QA device over other diode array alternatives, such as ArcCHECK. It allows for measurement at more realistic depths, thanks to the central location of the diodes housed within a thick casing of water-equivalent media. By comparison, the ArcCHECK array provides what is essentially a fluence-based measurement, with only a single point-dose taken at a realistic depth with the ionization chamber insert. MapCHECK is also able to attain a higher resolution than ArcCHECK (0.5 cm vs 1 cm resolution, since the helical arrangement of ArcCHECK's diodes prevents simple overlapping of measurements in the way that MapCHECK does).

4.4 Analysis of MapPHAN measurements

Kim *et al.* performed a sensitivity analysis of the gamma index method to MLC positioning errors in SABR [65]. They found that 2%/1 mm, with passing rates of 80% and 90%, was the most suitable gamma criteria for PSQA of VMAT SABR (a closely related technique to conformal arc SABR), out of the gamma criteria of 1%/1 mm, 1.5%/1.5 mm, 1%/2 mm, 2%/1 mm and 2%/2 mm. The results obtained here appear to confirm that 2%/1 mm with a threshold between 80% and 90% is an appropriate set of criteria to use, especially if one wishes to improve detection of OAR threshold failures. However, it would be useful to perform a more detailed analysis of the 5%/1 mm criteria with PTV results included, in order to further confirm its suitability for gamma indexing. Despite the lack of PTV analysis in this thesis, the detection accuracy of MapCHECK in MapPHAN to introduced errors was observed to be similar to that seen in the literature [63–68]. However, it is difficult to compare the ROC curve results obtained here with those obtained in the literature, due to the plethora of different devices and error types that were introduced, as well as the variation in gamma criteria between the two.

As it stands, gamma analysis has certain limitations when used as a metric for PSQA. Multiple papers have criticized the ambiguity of the test. Depuydt *et al.* [95] highlighted the potential for errors in the calculation for high dose gradient regions, especially when relative dose measurements are used. In their paper, the authors highlight the potential for certain errors to be missed due to a

lack of sensitivity to regions where a high density of points fail to meet the gamma criteria, while the rest of the plan passes. If one were to improve the gamma algorithm to highlight particular regions of interest where the potential for failure is greatest, then this would significantly improve the algorithm's QA benefits.

Several ROC analyses highlight the fact that the choice of parameters has a significant effect on the rate of failure of plans. For the centres that do optimise these criteria, the optimal rate of plan failure is commonly accepted as being 3%, regardless of the machine tolerances which are set in place for each centre, meaning that certain centres will be stricter in their evaluation of individual plans than others. For example, McKenzie *et al.* [96], during their substantive ROC curve analysis of different IMRT PSQA techniques, referenced Dong *et al.* [97] for the prevalence of an unacceptable plan - a study in which 751 plans were reviewed with ion chamber measurements. 97% of the IMRT monitor unit calculations agreed to within 3.5% of the ion chamber measurements for the delivered plans, in one of the largest studies of its kind that used a single TPS. However, this could be considered a weakness of using the study results to infer the proper cutoff for PSQA pass rates, as it was not a meta-analysis of multiple TPS's, departments, and measurement techniques, although it did sample plans from multiple treatment sites.

Additionally, the selection of optimal gamma analysis criteria based on ROC analysis will always introduce some arbitrary bias towards either a more conservative or more liberal approach to QA. Indeed, the authors of this paper found that determining the optimal gamma criteria based on the 3% threshold suggested by Dong *et al.* was essentially tantamount to performing a cost-weighting exercise, in which the cost of misclassifying an unacceptable plan as acceptable was $1/16^{\text{th}}$ of the cost of misclassifying an acceptable plan as unacceptable [96]. The final result of this was increased prioritization of efficiency over safety, since this cost analysis meant that passing an unacceptable plan carried less risk. This goes against the standard ethos of the medical physicist, whose clinical orientation is typically more conservative in the interest of ensuring absolute safety for patients, where possible. However, the unfortunate reality is that a financial analysis of any therapeutic system must set a limit on the minimal permissible efficiency of the system. Since it is much easier to determine the cost of failing an acceptable plan (in terms of equipment and man-hours) than to determine the cost of increased risk of passing an unacceptable plan, a slight bias towards efficiency of QA is typically introduced into QA optimization [96].

A significant disadvantage of 2D gamma analysis with MapPHAN is the lack of tissue-specific information it provides. SNC Patient software does not provide users with the ability to export the measured dose distribution, which would allow it to be overlaid on the CT slice from which it was originally taken. This could be done to obtain a visual comparison of the measured and calculated dose distributions with the CT dataset as a visual reference. This is not typically done, since it would only provide data for one slice in the case of film and MapCHECK. However, this is a limitation when trying to compare OAR and PTV doses between the measured and calculated treatment plans.

Granted, it would be virtually impossible to compare maximum point doses using a single slice of dosimetric data (highlighting a limitation of PSQA in general), but by taking the ratio of the measured and expected doses in a region aligned with a particular OAR on the CT dataset, one could infer the expected measured dose at other regions in the organ. The main limitation of this technique would be the rapid increase in uncertainty of interpolation accuracy as you move further away from the plane of dose measurement. Hence as it stands, gamma analysis provides no tissue-specific information, and is essentially a technique that focusses on PTV coverage, with very little consideration given to OAR doses. As such, hot and cold spots in the delivered plan may be ignored when the gamma pass rate is acceptable, especially when the pass rate is well within tolerance levels, despite the possibility of potential OAR overdosing.

Implications of superior and inferior isocentre shift results

One unexpected result from Table 3.3 was the difference observed between superior and inferior isocentre shift gamma results, as well as the difference between these shifted results and the gamma results for the original plans. From a beam's eye view perspective, the MapPHAN is an identical shape irrespective of couch position when the couch is shifted 1 mm in the inferior or superior direction. Granted, the diode position relative to the source will change when this occurs, but in theory, this should only become an issue when no shifts are included in SNC Patient. The results in Table 3.3 appear to confirm this. For example, consider the gamma results for the 1%/1 mm criteria with the phantom in the sagittal orientation. The difference between the mean gamma pass rate for merged results with isocentre shifts superior and inferior, when compared to the mean gamma pass rate for the original plans, is statistically insignificant (90.3±2.7, 90.3±3.9, and 88.8±2.5, respectively). This changes when looking at the same results, but with no SNC shifts applied

(87.8±3.8, 72.7±7.5, and 81.0±9.2, respectively). Note that all error bars were calculated at the 1 SD level. For the second case, there is a statistically significant shift between the gamma pass rates obtained for superior and inferior shifts. This can be attributed to the misalignment between the dose distributions due to the shifts, which is going uncorrected without the application of automatic shifts in SNC Patient. Similar discrepancies can be viewed throughout Table 3.3 for other criteria, although in general, the discrepancies are smaller for different shift directions. This can be explained by the fact that the phantom geometry changes relative to the source for all shift directions other than superior or inferior shifts; therefore, the dose distribution will also change slightly as a consequence, regardless of whether automatic shifts are included in SNC Patient.

The shifts performed by SNC were recorded throughout the experiment. The largest automatically-performed shift observed was (2 mm, 2 mm) in the x and y directions. This magnitude of positioning error would be remarkable if it were caused by the physicist setting up the phantom for QA. There were several shifts of (2 mm, 1 mm) also observed, along with (1 mm, 1 mm) shifts, most of which could be partly attributed to the isocentre shifts being introduced to the plans. The magnitude of these shifts seems to indicate that SNC Patient is overcorrecting for shifted dosimetric profiles. One option for increasing the sensitivity of QA to geometric errors would be to limit the maximum shift magnitude, perhaps to a vector of magnitude 2 mm in any direction. This would be done in the expectation that the QA physicist is unlikely to set up a phantom with positional error greater than this magnitude. Any machine errors of this magnitude or greater should also be detected during regular linac QA, in theory, so this should hopefully not factor significantly into the analysis.

The manual provided by Sun Nuclear for MapPHAN states the following: 'MapCHECK 2 with MapPHAN-MC2 results indicate that 1 mm positioning errors are detectable, again demonstrating the QA advantage of the high resolution diode detector.' No further information is provided regarding this statement. The results of this thesis are in clear opposition to this, especially when automatic software corrections are introduced in SNC Patient, despite the fact that these results are OAR specific. Stroom *et al.* [98] performed a robustness analysis on VMAT dose distributions with introduced isocentre shift errors. Their study found that isocentre shifts as large as 3 mm only had modest impacts on plan quality, with shifts of at least 1 - 2 mm being required to observe significant dosimetric disparities in the PTV. The study methodology was similar to that performed here,

suggesting that the errors introduced in this thesis were too small to provide data of clinical significance. However, this does call into question the validity of the statement provided by Sun Nuclear.

In summary, the evidence suggests that restrictions should be placed on shifts applied in SNC Patient for gamma analysis. The inclusion of shifts has the effect of significantly decreasing the detection of isocentre shift errors, and similar geometrical shifts. This is particularly the case for superior and inferior shifts, which become virtually impossible to detect due to the identical phantom geometry observed from a BEV perspective. This issue is unlikely to be unique to MapPHAN, as there are several phantoms that appear symmetrical along the superior-inferior axis. Additionally, the shifts used in this experiment, when applied, were significant, and were often random, suggesting that SNC Patient was overcorrecting for shifted dosimetric profiles. While further tests would need to be performed on other phantoms to confirm the results obtained here, it appears likely that restriction of automated shifts would significantly improve QA sensitivity and specificity to introduced geometrical errors.

4.4.1 Inter-group dosimetric error

A potential weakness of this study was the fact that measurements had to be taken on multiple days, due to time constraints. For most of the subjects, isocentre shift results were taken at a different time to the MLC and original results. Although dose was calibrated daily in order to negate the effect of variation in output of the linac, it is probable that taking measurements on different days had a significant effect on the magnitude of the dosimetric disparity between the planned and calculated dose distributions. Because of this, any trends observed in the central diode disparity should be isolated to the days on which they were taken. The validity of dose calibration is also reliant on the assumption that the central diode reading is a consistently accurate summary of the diode readings for the rest of the array. If the central diode itself exhibited any non-linearity over time, this would have shifted the dose calibration for the entire array. This sort of discrepancy would be difficult to detect without performing regular measurements comparing individual diodes on the array.

An example of the issues with performing dose calibrations on different days is as follows. If one treats the MLC shift and isocentre shift results as separate measurement groups, the central diode

disparity in the MLC results for Patient 2 may be compared directly to the disparity measured in the original plan delivery (as these results were measured on the same day), but cannot be directly compared to the disparity for isocentre results taken on a different day, as the variation in dose calibration will have a significant effect on the inter-group dosimetric comparison. This also makes the assumption that output fluctuations over the course of a day are relatively negligible.

4.4.2 Comparison of merged and unmerged MapCHECK measurements

A comparison of the merged and unmerged MapCHECK results indicate that the gamma pass rate is not significantly affected by increasing the resolution of measurement, with each combination of gamma criteria and level of introduced error exhibiting overlap in their respective error ranges. This implies that the merging of MapCHECK measurements is unnecessary for the QA of conformal arc SABR. However, the merging of measurements for SABR plans is recommended, in order to obtain more measurement points in high dose gradient regions. This in turn will make it easier to diagnose the cause of QA failure if and when it arises.

4.5 ROC curve analysis

Greiner *et al.* [59] proposed a generic set of guidelines for interpreting the AUC for ROC curves, using the following categories:

- AUC = 0.5 indicates that the test is non-informative;
- $0.5 < \text{AUC} \leq 0.7$ indicates a less accurate test;
- $0.7 < \text{AUC} \leq 0.9$ indicates a moderately accurate test;
- $0.9 < \text{AUC} < 1.0$ indicates a highly accurate test;
- AUC = 1.0 indicates that the test is perfect.

It should be remembered that only one or two individual classifiers will typically be utilised for gamma analysis - i.e. individual points on the curves will be selected for use in PSQA. Hence, in this case, the AUC is only a general indication of the accuracy of gamma criteria over the entire range of possible thresholds. However, the AUC remains useful for determining which gamma criteria should be selected for use in PSQA. The requisite threshold can be chosen after the gamma criteria themselves have been selected, using the Youden index or a similar method. In this case, based purely on the AUC values in Table 3.4, it appears that 5%/1 mm is the superior metric to use, although there is probably not enough data to make a statistically significant case for this. More realistically, one could suggest that a combination of 5%/1 mm and 1%/1 mm or 2%/1 mm criteria be used to generate an accurate assessment of the validity of QA plans.

One key difference between the curves generated for MLC and isocentre shift errors is the distribution of specific thresholds in ROC space for each set of gamma criteria. In general, the visible thresholds (95, 90, 85, 80, 75, 70, 65 and 60) are shifted farther towards the North-East corner of ROC space for the isocentre shift results. The effect can be seen most clearly when comparing the results for 1%/1 mm and 2%/1 mm, between the MLC and isocentre shift charts. This can be attributed to the introduction of MLC shifts having a greater dosimetric impact than the introduction of 1 mm isocentre shifts, in general. The plans that were most affected dosimetrically were in turn more likely to breach an OAR tolerance in Raystation. Since these plans contributed to the 'Positives' class as designated in Figure 2.6, they will have also contributed to the density of data in the centre of the curve. This has the benefit of increasing the positioning accuracy of the classifiers that are most clinically relevant, which are, ideally, closest to the point (0,1) in ROC space.

Another point of note is the variation in ROC data between orientations, for each respective set of gamma criteria. For instance, the difference between the curves generated in coronal and sagittal orientations for the 1%/1 mm criteria is significant, with the curve extending well below the line of $y = x$ for the coronal orientation - a trend that is not replicated in the sagittal orientation. Aside from the aforementioned statistical variation in these trends, there are other potential reasons for this discrepancy. The anisotropic nature of the dose distribution in Figure 3.2 means that, in theory, certain orientations of the phantom will be advantageous in detecting dose fluctuations in a particular axis.

For example, if the isocentre is shifted in the left-right direction on the patient couch, then placing

the phantom in the coronal orientation will, in theory, allow the physicist to detect significant dosimetric changes in this plane. (This assumes no shifts have been applied in SNC Patient software.) However, placing the phantom in the sagittal orientation, with the simulated target volume centred on the array, will prevent the detection of these sorts of fluctuations, as there is a relatively wide range of positions in which the dose distribution will appear virtually identical on the detector plane. This highlights the importance of performing QA with the detector array in multiple axes, as well as multiple plane positions if possible.

4.5.1 Accuracy of ROC curve predictions

The Area Under Curve metric (AUC) has previously been criticized as a comparison metric, since its accuracy can be misinterpreted when curves cross in ROC space. Hand [99] has declared that it uses 'different misclassification cost distributions for different classifiers. This means that using the AUC is equivalent to using different metrics to evaluate different classification rules'. In this thesis, the AUC was not an ideal metric for the analysis of gamma criteria, since the aim was to identify optimal discrete classifiers - in other words, to identify the optimal pass rate threshold that should be used for PSQA, in addition to the optimal gamma criteria that were used to generate the pass rates themselves. Granted, it is useful to know which gamma criteria perform best along the spectrum of discrete classifier points, but most of this information is made redundant as soon as one discrete classifier is chosen by a clinic. This is a serious limitation of gamma analysis, and is one reason why clinics should choose several gamma criteria for assessing plans. On the other hand, the AUC could potentially be useful for comparing the diagnostic abilities of different phantoms, for instance MapCHECK and ArcCHECK.

Flach and Wu [61] have posited a few options for repairing ROC curves with significant 'concavities' that extend down towards the point (1,0) in ROC space, such as those seen in Figure 3.15. One can repair these sections by simply mirroring the points in ROC space across the line from (0,0) to (1,1), which would improve a worse-than-random prediction model to a better-than-random model. Even by selecting diagonals between visibly optimal classifiers on the original ROC curve, and mirroring individual concavities across these lines, one can improve minor concavities in regions where the classifier model is performing poorly [61]. This is based on the assumption that the classifier being modified has useful information about the data but is applying it incorrectly. However, this begs the

question: can one be sure that the data is being incorrectly applied, rather than simply being worse than random by chance? Since the ROC curve is ultimately a statistical measure, one can never be absolutely certain. This must be considered in the case of isocentre shifts, where the curves are so far removed from the diagonal 'guessing strategy' that they seem quite implausible. A second check of these results confirms their validity for these measurements, however. One would expect that additional measurements would prove these results to be statistical aberrations, with the final result being much closer to the line of random guessing. However, it cannot be denied that the sensitivity of the phantom to isocentre shifts is less than desirable, as it stands.

It is difficult to verify how accurate the ROC curves are. Ideally a well-sampled ROC curve should appear as smooth as possible, in a manner analogous to that of a normal distribution with a high sample number. For ROC model specifiers that don't provide a continuous prediction spectrum, any vertical or horizontal lines almost certainly indicate a low sample number. This is because ROC curve data is generated using a decision tree method, in which you can only generate a discrete number of groups. The lack of sufficient data to generate statistically significant curves is a problem if one wants to use these curves in a prescriptive fashion - for instance, if one wishes to determine the optimal gamma criteria for a particular QA test. For the isocentre results obtained in this thesis, the sample size is far too low to make anything other than a descriptive assessment of the accuracy of certain gamma metrics, whereas for the MLC results, the sample size is probably high enough to warrant a more precise, quantitative analysis of superior gamma metrics.

Correlation between gamma pass rates and AUC

Figure 4.1 shows the AUC as a function of mean gamma pass rate, for all levels of introduced errors, based on Figure 3.17. The Pearson and Spearman correlation coefficients were calculated for each dataset, as well as for the combined dataset of both error types, as seen in Table 4.1.

From the coefficients calculated, it can be concluded that there is no clear linear correlation between AUC and mean gamma pass rate for individual tests. However, there are several corollaries to this. Due to the low number of datapoints, a calculation of Pearson's or Spearman's correlation coefficients would not provide a statistically significant assessment of the degree of correlation for either dataset. The high p-values in Table 4.1 confirm this lack of statistical significance. In addition, the combination of MLC and isocentre errors appears to form a quadratic trend, which

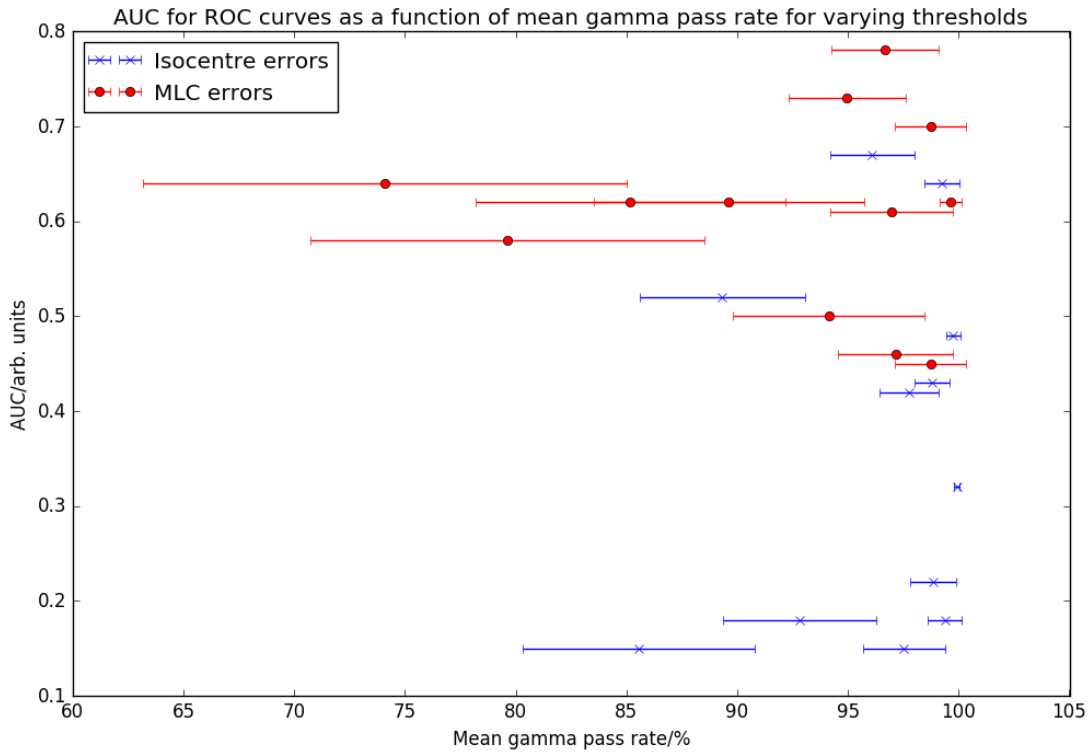


Figure 4.1: AUC as a function of mean gamma pass rate, for all levels of introduced MLC and isocentre shift errors.

cannot be fairly assessed by either Pearson's or Spearman's correlation coefficients as this is not a monotonically increasing function. Hence a larger dataset is required for an accurate assessment of any potential trends in either the MLC or isocentre shift error data, although an amalgamation of the two datasets is theoretically possible. If this analysis were completed, it could potentially be useful as a method for determining which mean gamma pass rate produces the greatest sensitivity and specificity. This could be used as evidence for the selection of an optimal set of gamma criteria, from which an ideal gamma threshold could be chosen via the Youden index.

Table 4.1: Pearson and Spearman correlation coefficients for datasets found in Figure 4.1, with associated p-values.

	Error type					
	MLC		Isocentre		Combined	
	Coefficient	p-value	Coefficient	p-value	Coefficient	p-value
Pearson	-0.05	0.87	0.20	0.54	-0.15	0.48
Spearman	-0.10	0.76	0.16	0.61	-0.22	0.31

4.6 Future work

This thesis has identified several potential areas of future work. In particular, the scope of the error analyses performed in this thesis could be expanded significantly. The ROC analysis of MapPHAN also needs substantial further investigation. These topics, and others, are discussed in the subsections below.

4.6.1 Determination of optimal gamma thresholds

The ROC curves generated in this thesis are useful, in that they provide a general indication of the best gamma criteria for PSQA of conformal arc plans. Generally, the performance of a single classifier can be quantified using the Youden index, which provides advantages over the AUC in that it is more specific to individual classifiers.

There are certain issues with this technique. Firstly, if one assumes that the original plan was delivered perfectly, then the Youden index is being calculated purely based on MLC errors. This is an unfair assessment of the high number of potential variations that could cause a QA plan to fail. The number of datapoints is also an issue, especially for the isocentre shift results. Both of these issues contribute to a further problem, which is the fact that the unbiased Youden index does not take into account the preferred sensitivity and specificity that is required for QA. In the case of radiotherapy, the cost of a false negative is high, as this could lead to a mistreatment of the patient. However, this must be balanced to a certain extent by economic factors, such as the required patient throughput for the centre. Hence using a particular set of criteria based purely on the Youden index would probably not be the best solution for any particular centre. However, it could be used as a potential starting point for further optimization.

It is normally the case that two sets of gamma criteria are used in tandem to detect treatment errors. Based solely on a visual estimate of the Youden index calculated for MLC results in both coronal and sagittal orientations, an appropriate combination of criteria for clinical use would be a 90% threshold with 2%/1 mm criteria, and a 95% threshold with 5%/1 mm criteria. This could be used as a starting point for QA, with further optimization possible around these criteria, depending on the specific needs of individual departments. This could be performed at multiple centres in order

to obtain a broader dataset for analysis of the optimal criteria to use for QA. The choice of 1 mm criteria is already appropriate for the high geometrical precision required of SABR treatments, which is a useful result.

4.6.2 Relevance of analysis to other techniques

The lung was chosen as the region of interest for this study, as this was the only site at Auckland Hospital being treated with SABR at the start of this project. This choice of location could be considered a hinderance, as the number of OARs surrounding a PTV in this region is relatively low compared to other regions of the body. In addition, the significant volume of low-density lung tissue adjacent to the treatment region has a significant influence on the shape of dose distributions, which is a unique feature of these treatments that isn't replicated elsewhere in the body. Further studies could investigate a variety of combinations of treatment sites and techniques, in order to determine to what extent the conformal arc plans are representative of other arc therapies, in terms of plan QA. In particular, an analysis of head and neck plans, in which there is a high density of OARs in close proximity to the PTV, would provide a significant addition to the *in silico* analysis provided here. Lung SABR plans in general have comparatively fewer OARs to analyse, making it difficult to obtain a high density of data. However, if further work were to be done on lung SABR plans, a focus on midline targets would be preferred in order to further quantify the effect of errors on heart and spine doses.

VMAT has been shown to be a superior treatment to both 3D-CRT and DCAT for the treatment of lung tumours using SABR, as it provides superior dose sparing to OARs [100]. As more departments shift from DCAT and 3D-CRT to VMAT for delivery of SABR, the QA methods used to analyse this technique will come under increased scrutiny. Repeating these experiments with patients treated using VMAT would produce markedly different trends to those seen in Figures 3.5 to 3.8. Field size is generally smaller in VMAT plans than in DCAT plans, with MLC leaves impinging on the treatment volume itself, from a beam's eye perspective. This means that introduced MLC errors would generate more significant dosimetric fluctuations in the centre of the PTV rather than around the periphery. This would likely reduce the maximum fluctuations in OAR doses seen in this thesis for introduced MLC shifts. However, the higher field perimeter to field size ratio will increase the significance of dosimetric fluctuations inside the PTV. It would be worthwhile to quantify this effect.

A similar trend to that seen in this thesis would probably be observed for the isocentre shift results, although the overall fluctuations would again be less marked than for the introduced MLC errors.

Due to the general shift from DCAT to VMAT SABR in clinics worldwide, the demand for high-resolution VMAT QA solutions is increasing. The experiments performed in this thesis prove that MapCHECK in MapPHAN can provide sufficient resolution for high dose gradient measurements. However, the phantom still needs to be validated for VMAT measurements, which can generate sharper dose gradients than conformal arc treatments. The additional modulation of gantry speed and dose rate, along with the increase in MLC modulation in terms of mean leaf velocity, could prove to have significant effects on the sensitivity and specificity of the phantom. In particular, the increased potential for interplay effect that MLC modulation in VMAT generates could have a significant effect on these results.

However, it should be noted that a number of departments plan their SABR lung VMAT plans with similar MLC field shapes and control point modulation to that seen in DCAT. This is done to reduce the effects of interplay between breathing motion and MLC aperture position, which can be the source of significant dosimetric uncertainties in the PTV [98, 101]. For departments with SABR VMAT treatments that are planned in this way, the results in this thesis are likely to be relevant to their QA practices.

4.6.3 Dosimetric analysis

The introduction of class shift errors to treatment plans has the advantage of producing generally linear dosimetric trends that are directly quantifiable. However, there are disadvantages to this method. Realistically, one would expect to see MLC leaf errors with a random distribution of negative and positive position errors relative to their expected position, since each MLC leaf is calibrated and initialised individually. Additionally, conformal arc plans include a collimator twist of 10^{o} or more. With collimator set to 0^{o}, each leaf is aligned with the plane of gantry rotation, meaning that individual leaf errors will also be aligned in one plane. Introduction of a small collimator shift negates this effect. This means that the dosimetric effect of an MLC leaf misalignment will be spread out over a broader range of tissues, rather than aligned in the plane of the misaligned leaf. However, Depuydt *et al.* showed that even single leaf pair errors, where one pair

of leaves fails during treatment delivery, can be picked up during gamma analysis [95]. Therefore, the effect cannot be ignored entirely.

In particular, the effect of MLC class open errors are relevant to plans delivered on Elekta linacs. The design of the MLC controllers on Elekta linacs (Elekta AB, Stockholm, Sweden) is such that the leaf banks are calibrated using an individual reference leaf as a guide for an entire leaf bank. The optically read-out position of this reference leaf is connected to the determined leaf position via a linear calibration, using an offset and a gain parameter. These parameters are defined as 'major' and 'minor' values, whereby the major offset and gain parameters act on the entire MLC leaf bank, and the minor parameters act on individual leaves [102]. Therefore, any errors in the calibration of the reference leaf will propagate into the entire leaf bank. (This can be contrasted with other MLC's such as Varian's Millenium MLC system, in which each leaf is calibrated individually [103].) The introduction of a collimator twist to treatment plans will be unable to prevent this type of error from having dosimetric effects as significant as the ones seen in this thesis, assuming the errors are of the same magnitude. This should serve as justification for a rigorous QA routine that should be performed on a daily basis (if not a per-treatment basis) for all MLC systems calibrated in this way.

One way to physically quantify the true effects of MLC position errors would be to generate plans with MLC positions based on Dynalog files of actual treatments. These files contain a histogram of the true MLC leaf position errors generated during treatments on Varian linacs. This would provide more realistic dosimetric outcomes that could then be used for analysis of QA, although the dataset required for a test like this would need to be substantial in order to obtain ROC curves with sufficient statistical information.

4.6.4 Clinical implementation of MapPHAN

Only one phantom, with one type of dosimeter in a planar array structure, was analysed during this study, which focussed only on plans with relatively small PTVs and high dose gradients. The relevance of the results in this thesis to other QA methods is debatable. While it could be argued that the MapCHECK phantom in MapPHAN, once properly commissioned, is a decent representation of the type of results that could be obtained with any planar geometry-based QA system (such as film), there is limited evidence in the literature to confirm this. One example that

could be cited is the paper by McKenzie *et al.* [96], whose ROC analysis of multiple IMRT QA techniques revealed similarities in the sensitivity and specificity of several planar (and non-planar) QA techniques for multiple gamma analysis criteria, with techniques that included MapCHECK coupled with SNC Patient software, EDR2 film with I'mRT software, and ArcCHECK with SNC Patient software, which contains a helical diode array structure. However, this study was based on a limited patient dataset, and a meta-analysis of multiple QA techniques for an extensive range of treatment scenarios would provide a better indication of how accurately one phantom can represent other QA techniques.

It would be worthwhile to see if these ROC curve patterns re-emerged with other phantom geometries, such as the peripheral helical array geometry found in the ArcCHECK phantom. This would provide an indication of whether these patterns are phantom independent, a conclusion that can be tentatively derived from the comparison between merged and non-merged results in this study. This prediction would be made based on the assumption that the analysis is detector-resolution independent, as ArcCHECK has a minimum resolution of 1 cm spacing between diodes, even with merged measurements.

The use of 1 mm isocentre shifts, while clinically relevant as they represented the limits of machine QA tolerances at ADHB, limited the predictive accuracy of the ROC curves generated for MapPHAN. The introduction of greater shifts, perhaps between 2-3 mm, would significantly increase the density of data in the portion of the curve that contributes to the calculation of AUC. The magnitude of these introduced errors would make these results more relevant to techniques that don't deliver ablative doses, but the data obtained would provide more accurate insights into the value of using phantoms to detect positioning errors. In theory, it would also lend more credence to the idea that applying automatic shifts in SNC Patient (and similar software) significantly reduces the detection of isocentre shift errors, as the data obtained for this thesis are not entirely conclusive on this point.

Work is currently being done to develop the 'Edgeless' diode for radiotherapy [55]. This detector purportedly achieves an angular dependence of less than ± 2%. These results were measured with the diode housed in a cylindrical acrylic phantom, which was rotated with bidirectional accuracy of 0.25^{o} beneath a static 10 x 10 cm^2 6 MV X-ray beam delivered at G0 [55]. This technique eliminated any dose uncertainties introduced by rotation of the gantry. The diode is fabricated using an 'active

edge' technology developed by VTT Technical Research Centre of Finland Micro and Nanoelectronics, which uses ion implantation to activate the silicon wafer's lateral junctions [104], significantly reducing the diode's dependence for lateral beams (i.e. at G90 and G270). This manufacturing process is used in combination with 'drop in' packaging technology used by the Centre for Medical Radiation Physics [105]. The drop-in technology consists of a flexible carrier with no high-Z materials, from which the detector die hangs. The lack of high-Z inhomogeneities means that secondary electrons produced by the incoming X-ray beam can reach the active volume at an approximately equal fluence rate from all directions. If successfully commissioned, this type of diode would eliminate the necessity of approximating angular independence with density overrides. In the case of MapCHECK, this would allow the full extent of the array to be utilised for PSQA, without concern for increasing uncertainties towards the edges of the array.

Chapter 5

Conclusion

In this thesis, ROC curves were used to examine the efficacy of using MapPHAN for the QA of DCAT lung SABR plans. A particular emphasis was placed on the phantom's ability to detect dosimetric errors in normal tissues.

MapCHECK with MapPHAN is a viable alternative to film for performing QA on DCAT lung SABR plans. It provides sufficient resolution to measure in high dose-gradient regions, while providing measurements at a depth close to the typical position of the target volume. The phantom exhibits good sensitivity and specificity to introduced MLC errors, with results being comparable to those seen in the literature. The poor accuracy of the phantom for detection of introduced isocentre shift errors is expected, given that the magnitude of the errors we introduced (less than or equal to 1 mm) was low compared to the magnitude required for significant dosimetric effects (greater than 1 mm). However, restrictions could be placed on the shift magnitude automatically applied in SNC patient, or shifts removed entirely, in order to improve the sensitivity of the phantom to these errors. Increasing the resolution of the phantom by merging two measurements does not have a significant effect on the phantom's gamma pass rate (and hence sensitivity or specificity), but should be performed if the physicist requires more detectors in high dose gradient regions.

Counter to manufacturer's recommendations, the combination of MapCHECK housed in MapPHAN does require extensive corrections to account for the inherent angular dependencies of its diodes. Nonetheless, it can be commissioned to the point where absolute dosimetric measurements are possible for a restricted section of the diode array (within 4 cm laterally of the central diode

column). The application of variable density overrides to different sections of the phantom is a feasible option for attaining the requisite level of accuracy, although the densities will need to be optimised by each individual centre.

Based on the ROC curve analysis performed, a combination of 2%/1 mm with 90% threshold and 5%/1 mm with 95% threshold has been identified as an appropriate gamma analysis protocol for conformal arc SABR treatments. These thresholds should be further optimized according to the clinical and economical needs of each centre.

Chapter 6

Appendix

.1 Python code for class shift errors

The code shown below was used to introduce class shift errors to the original plans. These plans were then imported into Raystation with the edited RTPlan files, and recalculated.

```
# Author: Nathan Henry, 2017
# This script takes an RTPlan file, and introduces a class shift error to
    both MLC banks for each control point, before generating a new RTPlan
    file with the edits included.

from pydicom import dicomio
import os
import numpy as np
from sys import version_info

# This section tests for python version and gets input from user to
   determine filename

# Create boolean value for test that Python major version > 2
```

```
py3 = version_info[0] > 2

if py3:
  filename = input("Enter filename: ")
else:
  filename = raw_input("Enter filename: ")

# This section edits dicom tag

ds = dicomio.read_file(filename)

# This value represents the class leaf position error in mm for both MLC
    banks
increment = 0.2

# This section pulls out an array of MLC position values and edits them.
    It cycles through all possible beams and control points to select the
    maximum possible value for both, before editing all control points for
     each beam.

beam_num = 0
q = 0
while (q == 0):
    try:
        blah = ds[0x300a,0x00b0][beam_num][0x300a,0x0111]
        beam_num = beam_num + 1
    except:
        beam_num = beam_num - 1
        q = 1

print('Number of beams = ')
```

```
print(beam_num + 1)
for k in range(0,beam_num + 1):
    x = ds[0x300a,0x00b0][k][0x300a,0x0111][0][0x300a,0x011a][2][0x300a,0
        x011c]
    q = 0
    ctrl_point = 0
    while (q == 0):
        try:
            blay = ds[0x300a,0x00b0][k][0x300a,0x0111][ctrl_point][0x300a
                ,0x011a]
            ctrl_point = ctrl_point + 1
        except:
            ctrl_point = ctrl_point - 1
            q = 1

    print('Current beam number = ')
    print(k+1)
    print('Number of control points for this beam = ')
    print(ctrl_point + 1)

# First control point has extra 2 beam limiting devices - jaws

    print('\n')
    print('Beam number = ')
    print(k + 1)
    print('Control point =')
    print(1)
    print('Original MLC positions = ')
    print(ds[0x300a,0x00b0][k][0x300a,0x0111][0][0x300a,0x011a][2][0x300a
        ,0x011c])
    print('New MLC positions = ')
```

```
    for i in range(0,60):
        ds[0x300a,0x00b0][k][0x300a,0x0111][0][0x300a,0x011a][2][0x300a,0
            x011c].value[i] = ds[0x300a,0x00b0][k][0x300a,0x0111][0][0
            x300a,0x011a][2][0x300a,0x011c].value[i]-increment
    for j in range(60,120):
        ds[0x300a,0x00b0][k][0x300a,0x0111][0][0x300a,0x011a][2][0x300a,0
            x011c].value[j] = ds[0x300a,0x00b0][k][0x300a,0x0111][0][0
            x300a,0x011a][2][0x300a,0x011c].value[j]+increment

    print(ds[0x300a,0x00b0][k][0x300a,0x0111][0][0x300a,0x011a][2][0x300a
        ,0x011c])

# Insert new MLC positions into RTPlan file

    for m in range (1, ctrl_point +1):
        print('\n')
        print('Beam number =')
        print(k+1)
        print('Control point =')
        print(m+1)
        print('Original MLC positions = ')
        print(ds[0x300a,0x00b0][k][0x300a,0x0111][m][0x300a,0x011a][0][0
            x300a,0x011c])
        print('New MLC positions = ')
        for i in range(0,60):
            ds[0x300a,0x00b0][k][0x300a,0x0111][m][0x300a,0x011a][0][0
                x300a,0x011c].value[i] = ds[0x300a,0x00b0][k][0x300a,0
                x0111][m][0x300a,0x011a][0][0x300a,0x011c].value[i]-
                increment
        for j in range(60,120):
```

```
                ds[0x300a,0x00b0][k][0x300a,0x0111][m][0x300a,0x011a][0][0
                    x300a,0x011c].value[j] = ds[0x300a,0x00b0][k][0x300a,0
                    x0111][m][0x300a,0x011a][0][0x300a,0x011c].value[j]+
                    increment
            print(ds[0x300a,0x00b0][k][0x300a,0x0111][m][0x300a,0x011a][0][0
                x300a,0x011c])

# This section renames file

##new_filename = filename.replace(".dcm","_edited.dcm")
##
##print(new_filename)
##ds.save_as(new_filename)

# Alternatively, this section replaces old RTPlan file with edited file

ds.save_as(filename)

print("Done")
```

.2 Volumetric DVH parameter analysis of plans

The following figures show the increase in volume of OAR that would receive above their threshold dose (as seen in Table 2.1) as a result of introduced MLC errors. The gradients are similar to those used to generate Figure 3.3, except the DVH constraints used were volumetric rather than point-dose based. Dashed lines indicate the thresholds at which the associated volumetric OAR constraint would theoretically fail, if the potential MLC-induced volumetric errors were taken into account. Due to the low number of datapoints available for each of these OARs, any attempt at applying trends to the data would have been statistically irrelevant.

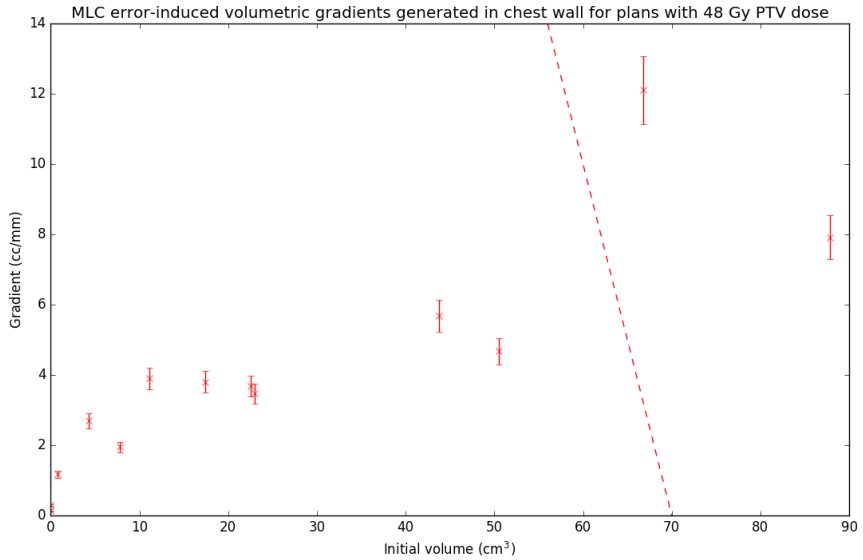


Figure 1: Volumetric gradients generated in chest wall from introduction of MLC errors to plans with a PTV dose of 48 Gy. Error bars quoted to within ± 2 SD.

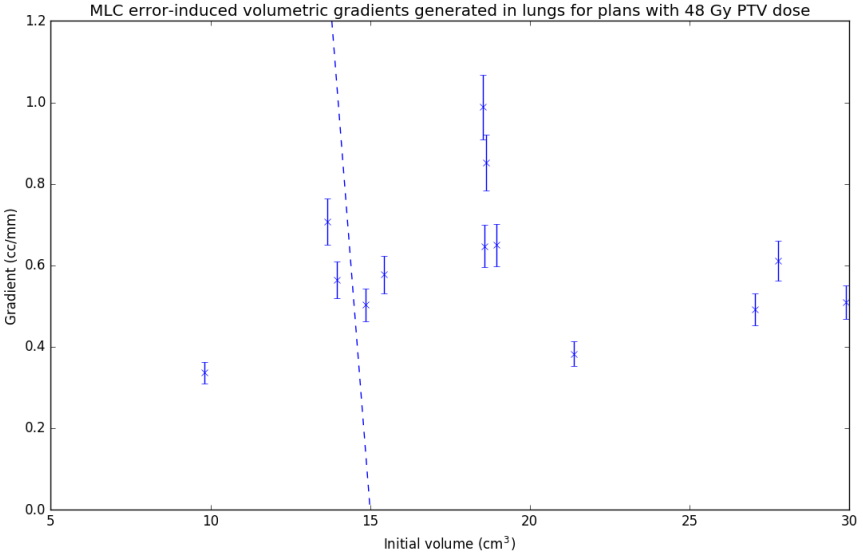


Figure 2: Volumetric gradients generated in lungs from introduction of MLC errors to plans with a PTV dose of 48 Gy. Error bars quoted to within ± 2 SD.

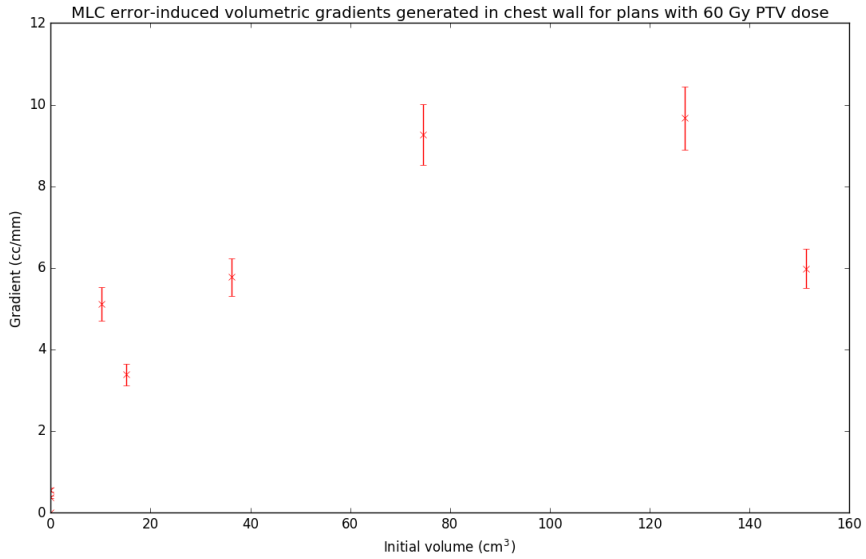


Figure 3: Volumetric gradients generated in chest wall from introduction of MLC errors to plans with a PTV dose of 60 Gy. Error bars quoted to within ± 2 SD.

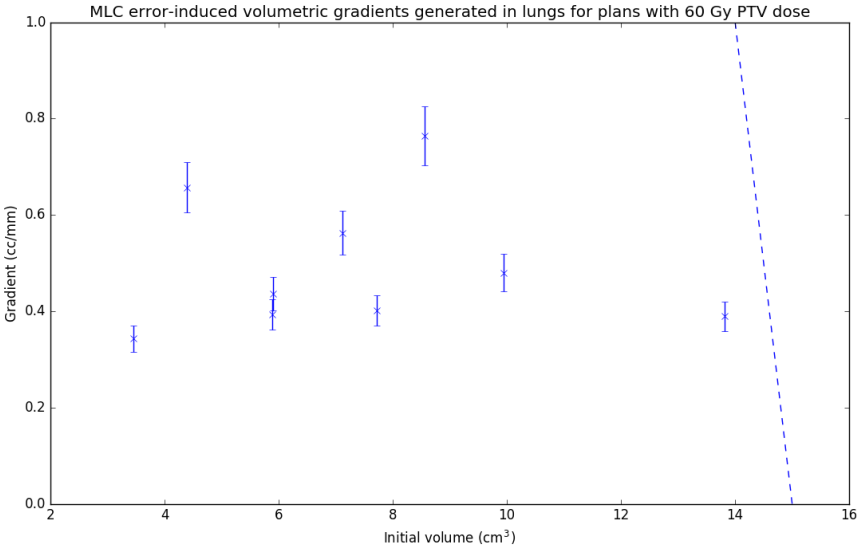


Figure 4: Volumetric gradients generated in lungs from introduction of MLC errors to plans with a PTV dose of 60 Gy. Error bars quoted to within ± 2 SD.

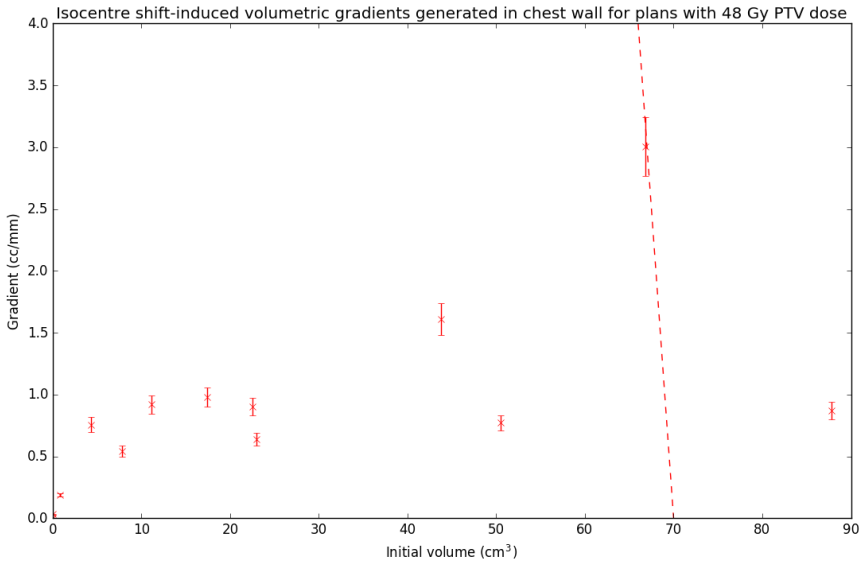


Figure 5: Volumetric gradients generated in chest wall from introduction of isocentre shift errors to plans with a PTV dose of 48 Gy. Error bars quoted to within ± 2 SD.

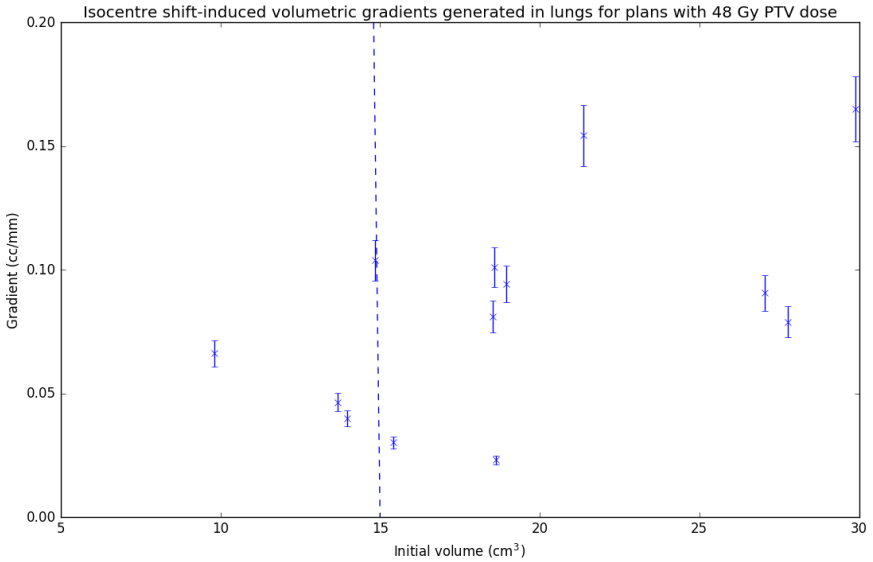


Figure 6: Volumetric gradients generated in lungs from introduction of isocentre shift errors to plans with a PTV dose of 48 Gy. Error bars quoted to within ± 2 SD.

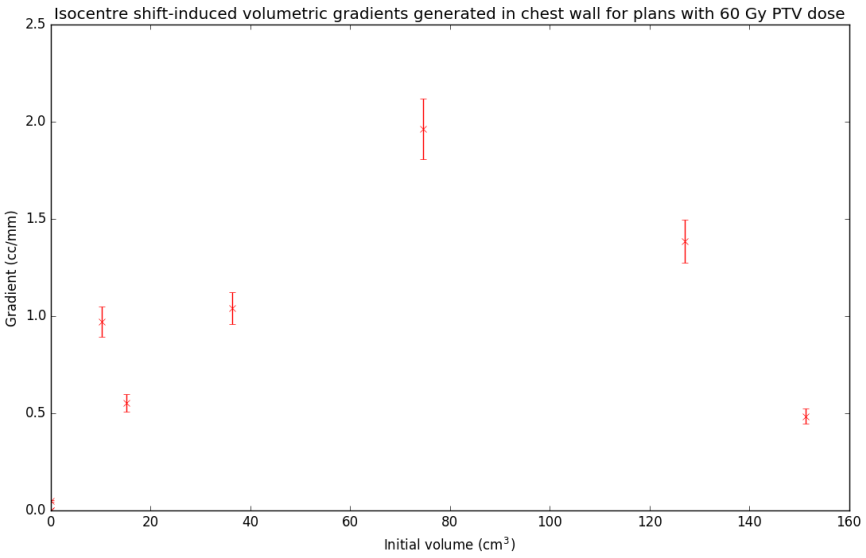


Figure 7: Volumetric gradients generated in chest wall from introduction of isocentre shift errors to plans with a PTV dose of 60 Gy. Error bars quoted to within ± 2 SD.

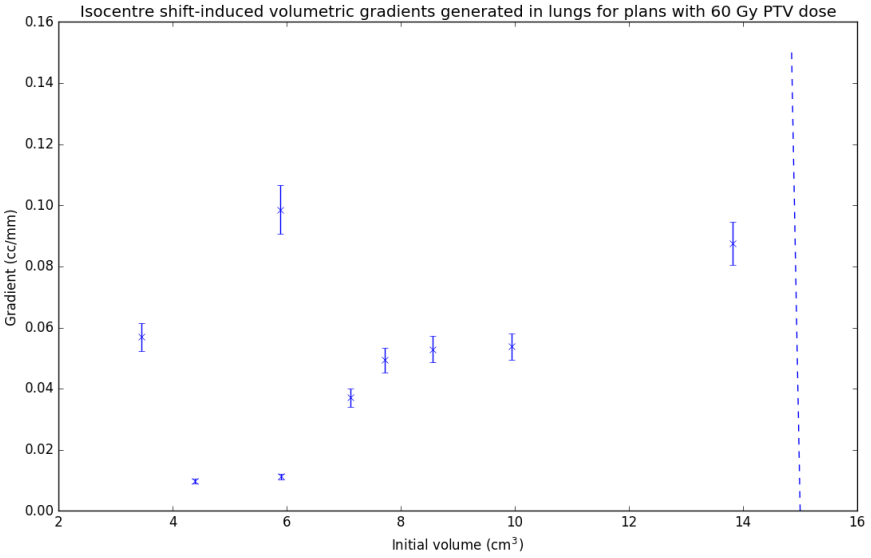


Figure 8: Volumetric gradients generated in lungs from introduction of isocentre shift errors to plans with a PTV dose of 60 Gy. Error bars quoted to within ± 2 SD.

.3 Angular dependency of MapCheck in MapPhan

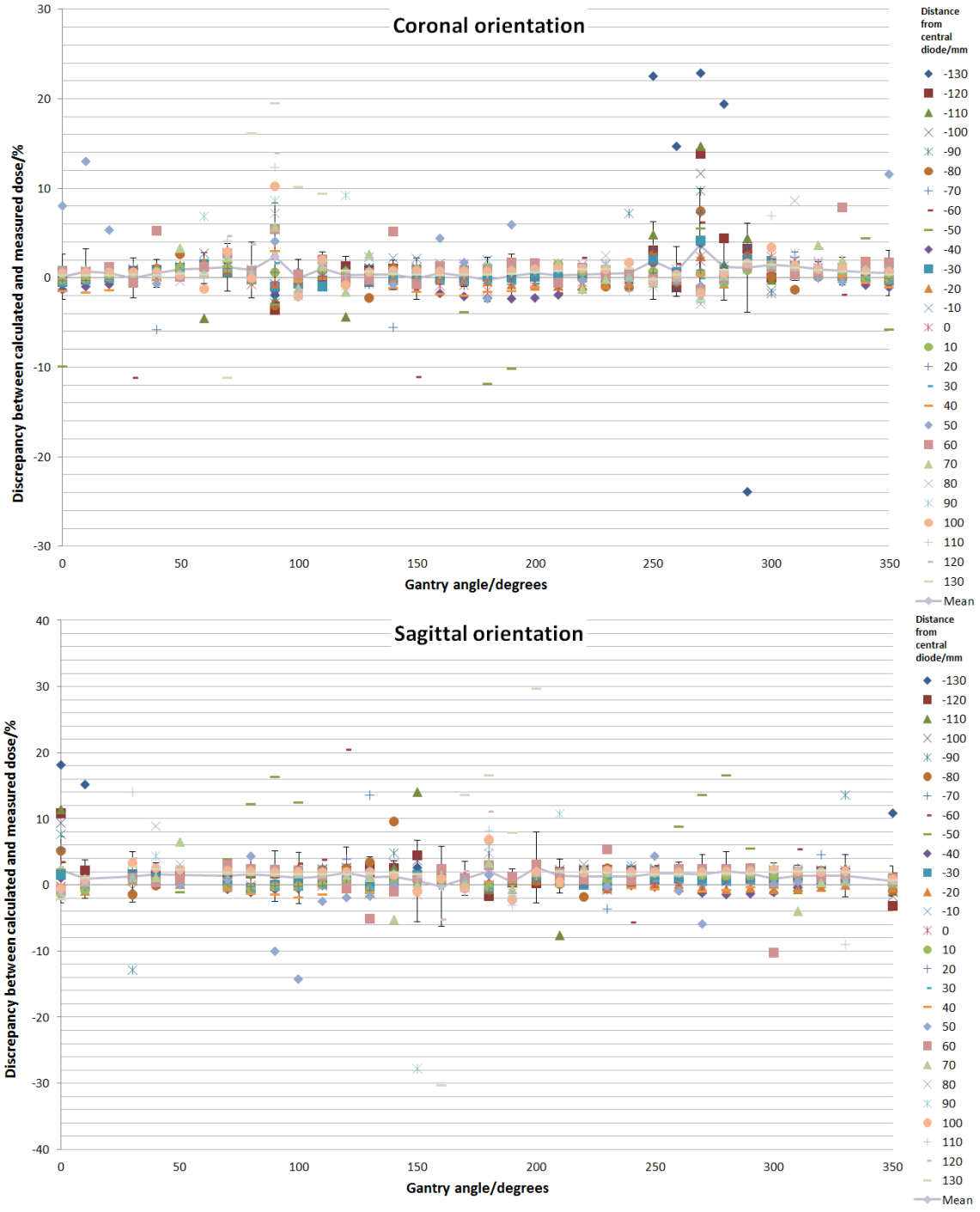
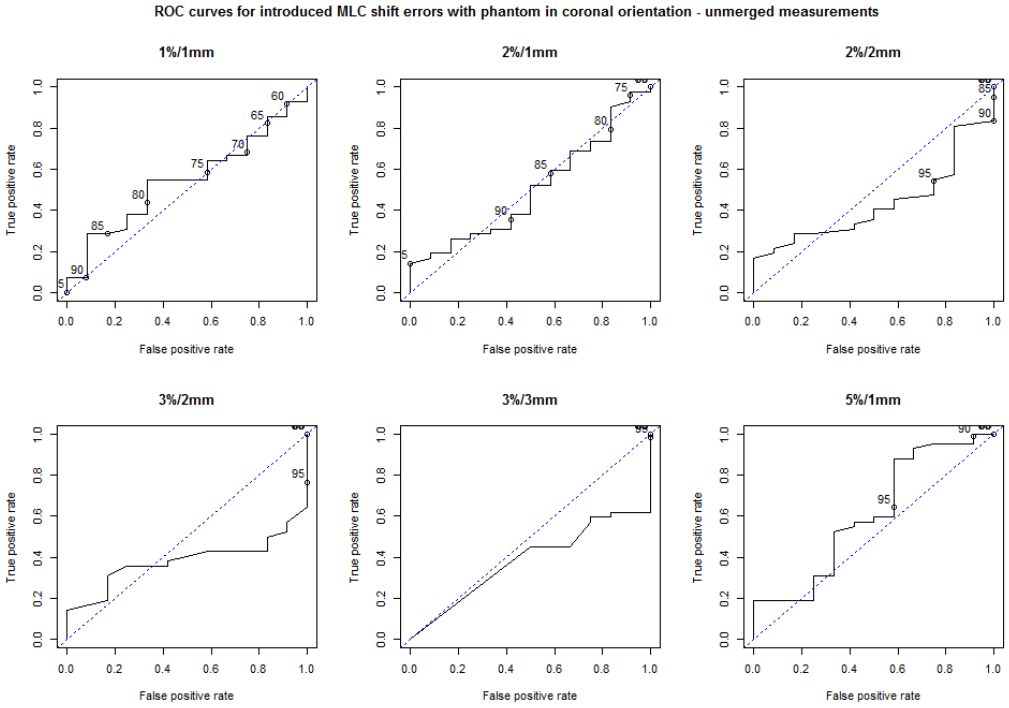
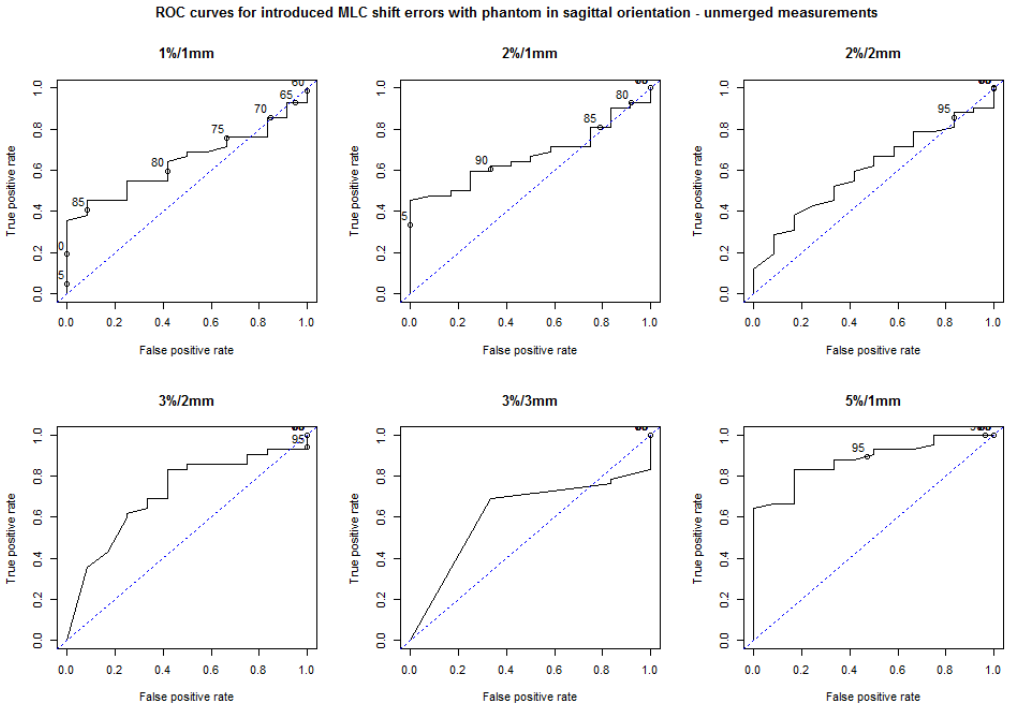


Figure 9: Ratio of calculated/measured dose at all diode positions along central transverse axis, for angular dependency measurements on MapCheck in coronal and sagittal orientations. Trend line represents mean discrepancy for each gantry angle. Error bars represent 1 standard deviation from the mean.

.4 ROC curves for unmerged MapPhan results

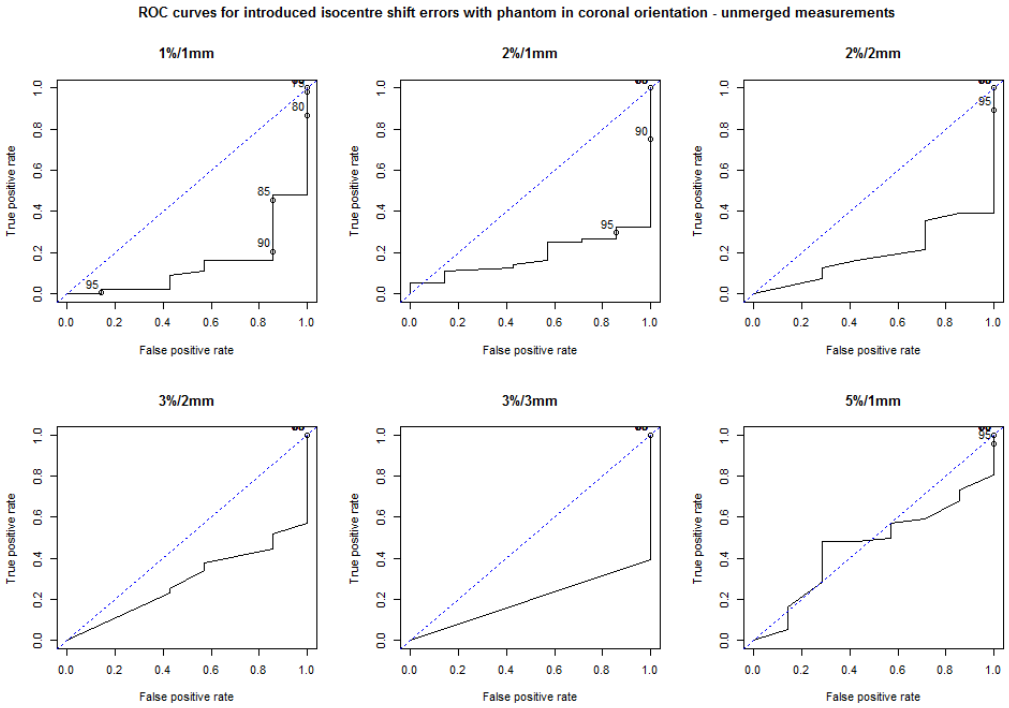


(a)

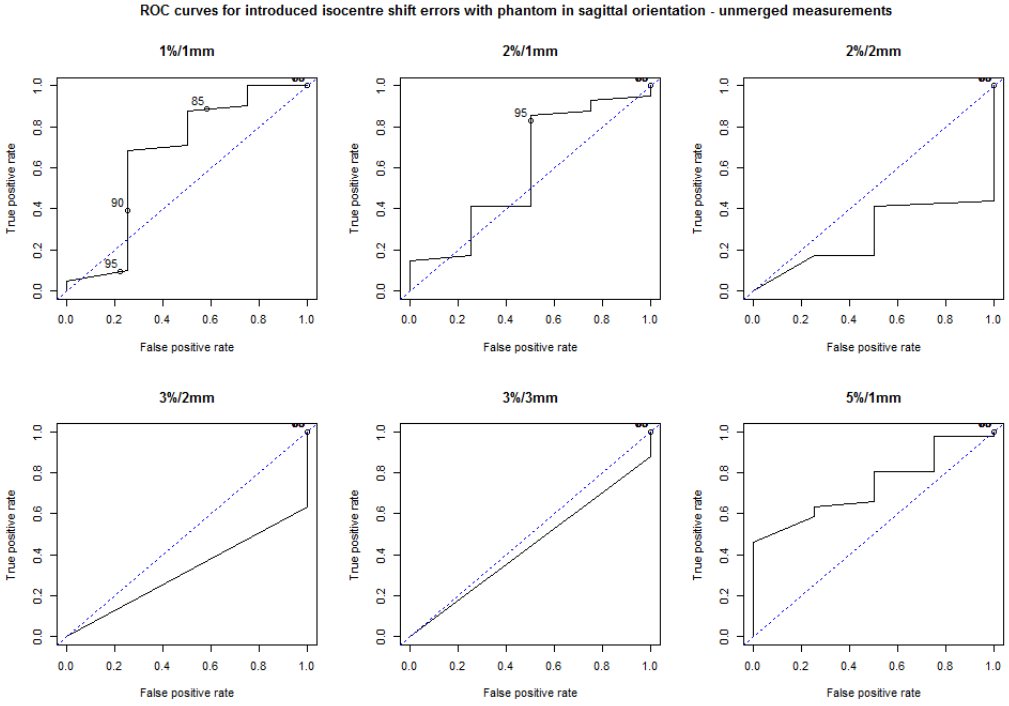


(b)

Figure 10: ROC curves generated from unmerged MapPhan results with multiple gamma analysis criteria, for plans with introduced MLC shift errors delivered to MapPhan in both coronal (top) and sagittal (bottom) orientations.



(a)



(b)

Figure 11: ROC curves generated from unmerged MapPhan results with multiple gamma analysis criteria, for plans with introduced isocentre shift errors delivered to MapPhan in both coronal (top) and sagittal (bottom) orientations.

Bibliography

[1] I. C. on Radiation Units and Measurements, *ICRU report 83 prescribing, recording, and reporting photon-beam intensity-modulated radiation therapy (IMRT)-Journal of the ICRU-vol 10 no 1 2010*. Oxford University Press, 2010.

[2] P. Rowshan Farzad *et al.*, "Improvement of epid-based techniques for dosimetry and investigation of linac mechanical performance in advanced radiotherapy," 2012.

[3] R. D. Evans, "Compton effect," in *Corpuscles and Radiation in Matter II/Korpuskeln und Strahlung in Materie II*, pp. 218–298, Springer, 1958.

[4] J. Malhotra, M. Malvezzi, E. Negri, C. La Vecchia, and P. Boffetta, "Risk factors for lung cancer worldwide," *European Respiratory Journal*, pp. ERJ–00359, 2016.

[5] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, and D. M. Parkin, "Estimates of worldwide burden of cancer in 2008: Globocan 2008," *International journal of cancer*, vol. 127, no. 12, pp. 2893–2917, 2010.

[6] R. Haynes, J. Pearce, and R. Barnett, "Cancer survival in new zealand: ethnic, social and geographical inequalities," *Social science & medicine*, vol. 67, no. 6, pp. 928–937, 2008.

[7] A. Jemal, F. Bray, M. M. Center, J. Ferlay, E. Ward, and D. Forman, "Global cancer statistics," *CA: a cancer journal for clinicians*, vol. 61, no. 2, pp. 69–90, 2011.

[8] S. Siva and B. J. Slotman, "Sabr for lung metastases–where is the evidence and what are we doing with it?," in *Seminars in Radiation Oncology*, Elsevier, 2017.

[9] D. S. Ettinger, W. Akerley, G. Bepler, M. G. Blum, A. Chang, R. T. Cheney, L. R. Chirieac, T. A. D'Amico, T. L. Demmy, A. K. P. Ganti, *et al.*, "Non–small cell lung cancer," *Journal of the national comprehensive cancer network*, vol. 8, no. 7, pp. 740–801, 2010.

[10] T.-Y. D. Cheng, S. M. Cramb, P. D. Baade, D. R. Youlden, C. Nwogu, and M. E. Reid, "The international epidemiology of lung cancer: latest trends, disparities, and tumor characteristics," *Journal of Thoracic Oncology*, vol. 11, no. 10, pp. 1653–1671, 2016.

[11] P. Koller, "Biological basis of radiotherapy," *Cancer*, vol. 5, no. Pt 9, pp. 28–53, 1959.

[12] J. Chavaudra and A. Bridier, "Definition of volumes in external radiotherapy: Icru reports 50 and 62," *Cancer radiotherapie: journal de la Societe francaise de radiotherapie oncologique*, vol. 5, no. 5, pp. 472–478, 2001.

[13] D. Greene and P. C. Williams, *Linear accelerators for radiation therapy.* CRC Press, 1997.

[14] C. Bhatt, I. Ahmad, M. Semwal, and K. Chufal, "Technological development, clinical application," *Quality Assurance and Dosimetric Validation of Volumetric Modulated Arc Therapy (VMAT): A Comprehensive Literature Review. J Nucl Med Radiat Ther S*, vol. 9, p. 2, 2018.

[15] J. Knipp and G. Uhlenbeck, "Emission of gamma radiation during the beta decay of nuclei," *Physica*, vol. 3, no. 6, pp. 425–439, 1936.

[16] F. M. Khan and J. P. Gibbons, *Khan's the physics of radiation therapy.* Lippincott Williams & Wilkins, 2014.

[17] L. Spencer and F. H. Attix, "A theory of cavity ionization," *Radiation Research*, vol. 3, no. 3, pp. 239–254, 1955.

[18] F. H. Attix, *Introduction to radiological physics and radiation dosimetry.* John Wiley & Sons, 2008.

[19] B. J. Gerbi and F. M. Khan, "Measurement of dose in the buildup region using fixed-separation plane-parallel ionization chambers," *Medical physics*, vol. 17, no. 1, pp. 17–26, 1990.

[20] L. Apipunyasopon, S. Srisatit, and N. Phaisangittisakul, "An investigation of the depth dose in the build-up region, and surface dose for a 6-mv therapeutic photon beam: Monte carlo simulation and measurements," *Journal of radiation research*, vol. 54, no. 2, pp. 374–382, 2012.

[21] N. R. Council *et al.*, *Health risks from exposure to low levels of ionizing radiation: BEIR VII phase 2*, vol. 7. National Academies Press, 2006.

[22] K. Chadwick and H. Leenhouts, "A molecular theory of cell survival," *Physics in Medicine & Biology*, vol. 18, no. 1, p. 78, 1973.

[23] E. J. Hall, A. J. Giaccia, *et al.*, *Radiobiology for the Radiologist*, vol. 6. Lippincott Williams & Wilkins Philadelphia:, 2006.

[24] M. Williams, J. Denekamp, and J. Fowler, "A review of α β ratios for experimental tumors: implications for clinical studies of altered fractionation," *International Journal of Radiation Oncology Biology Physics*, vol. 11, no. 1, pp. 87–96, 1985.

[25] J. T. Lyman, "Complication probability as assessed from dose-volume histograms," *Radiation Research*, vol. 104, no. 2s, pp. S13–S19, 1985.

[26] B. Warkentin, P. Stavrev, N. Stavreva, C. Field, and B. G. Fallone, "A tcp-ntcp estimation module using dvhs and known radiobiological models and parameter sets," *Journal of Applied Clinical Medical Physics*, vol. 5, no. 1, pp. 50–63, 2004.

[27] G. Mitchell, "The rationale for fractionation in radiotherapy.," *Clinical journal of oncology nursing*, vol. 17, no. 4, 2013.

[28] A. E. Nahum, "The radiobiology of hypofractionation," *Clinical Oncology*, vol. 27, no. 5, pp. 260–269, 2015.

[29] M. R. Folkert and R. D. Timmerman, "Stereotactic ablative body radiosurgery (sabr) or stereotactic body radiation therapy (sbrt)," *Advanced drug delivery reviews*, vol. 109, pp. 3–14, 2017.

[30] S. Dische and M. I. Saunders, "Continuous, hyperfractionated, accelerated radiotherapy (chart): an interim report upon late morbidity," *Radiotherapy and Oncology*, vol. 16, no. 1, pp. 65–72, 1989.

[31] M. I. Saunders and S. Dische, "Continuous, hyperfractionated, accelerated radiotherapy (chart) in non-small cell carcinoma of the bronchus," *International Journal of Radiation Oncology Biology Physics*, vol. 19, no. 5, pp. 1211–1215, 1990.

[32] F. Alongi, A. Fiorentino, and B. De Bari, "Sbrt and extreme hypofractionation: A new era in prostate cancer treatments?," *Reports of Practical Oncology & Radiotherapy*, vol. 20, no. 6, pp. 411–416, 2015.

[33] Y. Park, H. J. Kim, and A. R. Chang, "Predictors of chest wall toxicity after stereotactic ablative radiotherapy using real-time tumor tracking for lung tumors," *Radiation Oncology*, vol. 12, no. 1, p. 66, 2017.

[34] F. Soldà, M. Lodge, S. Ashley, A. Whitington, P. Goldstraw, and M. Brada, "Stereotactic radiotherapy (sabr) for the treatment of primary non-small cell lung cancer; systematic review and comparison with a surgical cohort," *Radiotherapy and Oncology*, vol. 109, no. 1, pp. 1–7, 2013.

[35] V. Verma, C. B. Simone, and W. Zhen, "Stereotactic radiotherapy for stage i small cell lung cancer," 2016.

[36] H. Chen, J. Laba, G. Boldt, C. Goodman, D. Palma, S. Senan, and A. Louie, "Stereotactic ablative radiotherapy versus surgery in early lung cancer: A meta-analysis of propensity score-adjusted comparative effectiveness studies," *International Journal of Radiation Oncology Biology Physics*, vol. 99, no. 2, p. E445, 2017.

[37] S. S. Lo, M. Foote, S. Siva, B. J. Slotman, B. S. Teh, M. Guckenberger, D. Tan, N. A. Mayr, and A. Sahgal, "Technical know-how in stereotactic ablative radiotherapy (sabr)," *Journal of medical radiation sciences*, vol. 63, no. 1, pp. 5–8, 2016.

[38] X. Qiao, O. Tullgren, I. Lax, F. Sirzén, and R. Lewensohn, "The role of radiotherapy in treatment of stage i non-small cell lung cancer," *Lung cancer*, vol. 41, no. 1, pp. 1–11, 2003.

[39] H. Onishi, H. Shirato, Y. Nagata, M. Hiraoka, M. Fujino, K. Gomi, Y. Niibe, K. Karasawa, K. Hayakawa, Y. Takai, *et al.*, "Hypofractionated stereotactic radiotherapy (hypofxsrt) for stage i non-small cell lung cancer: updated results of 257 patients in a japanese multi-institutional study," *Journal of thoracic oncology*, vol. 2, no. 7, pp. S94–S100, 2007.

[40] S. D. McGrath, M. M. Matuszak, D. Yan, L. L. Kestin, A. A. Martinez, and I. S. Grills, "Volumetric modulated arc therapy for delivery of hypofractionated stereotactic lung radiotherapy: A dosimetric and treatment efficiency analysis," *Radiotherapy and Oncology*, vol. 95, no. 2, pp. 153–157, 2010.

[41] B. S. Teh, S. Y. Woo, and E. B. Butler, "Intensity modulated radiation therapy (imrt): a new promising technology in radiation oncology," *The oncologist*, vol. 4, no. 6, pp. 433–442, 1999.

[42] M. Monz, K.-H. Küfer, T. R. Bortfeld, and C. Thieke, "Pareto navigationalgorithmic foundation of interactive multi-criteria imrt planning," *Physics in Medicine & Biology*, vol. 53, no. 4, p. 985, 2008.

[43] M. Zhang, S.-M. Zhou, and T. Qu, "What do we mean when we talk about the linac isocenter?," *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*, vol. 4, no. 03, p. 233, 2015.

[44] W. Du, S. Gao, X. Wang, and R. J. Kudchadker, "Quantifying the gantry sag on linear accelerators and introducing an mlc-based compensation strategy," *Medical physics*, vol. 39, no. 4, pp. 2156–2162, 2012.

[45] T. Gupta, J. Agarwal, S. Jain, R. Phurailatpam, S. Kannan, S. Ghosh-Laskar, V. Murthy, A. Budrukkar, K. Dinshaw, K. Prabhash, *et al.*, "Three-dimensional conformal radiotherapy (3d-crt) versus intensity modulated radiation therapy (imrt) in squamous cell carcinoma of the head and neck: a randomized controlled trial," *Radiotherapy and Oncology*, vol. 104, no. 3, pp. 343–348, 2012.

[46] C. L. Ong, W. F. Verbakel, J. P. Cuijpers, B. J. Slotman, F. J. Lagerwaard, and S. Senan, "Stereotactic radiotherapy for peripheral lung tumors: a comparison of volumetric modulated arc therapy with 3 other delivery techniques," *Radiotherapy and oncology*, vol. 97, no. 3, pp. 437–442, 2010.

[47] G. G. Zhang, L. Ku, T. J. Dilling, C. W. Stevens, R. R. Zhang, W. Li, and V. Feygelman, "Volumetric modulated arc planning for lung stereotactic body radiotherapy using conventional and unflattened photon beams: a dosimetric comparison with 3d technique," *Radiation oncology*, vol. 6, no. 1, p. 152, 2011.

[48] X. Y. Cedric, "Intensity modulated arc therapy with dynamic multi-leaf collimation," Oct. 6 1998. US Patent 5,818,902.

[49] S.-M. Chae, K. W. Lee, and S. H. Son, "Dosimetric impact of multileaf collimator leaf width according to sophisticated grade of technique in the imrt and vmat planning for pituitary adenoma lesion," *Oncotarget*, vol. 7, no. 47, p. 78119, 2016.

[50] L. E. Court, M. Wagar, D. Ionascu, R. Berbeco, and L. Chin, "Management of the interplay effect when using dynamic mlc sequences to treat moving targets," *Medical physics*, vol. 35, no. 5, pp. 1926–1931, 2008.

[51] C.-W. Cheng and I. J. Das, "Treatment plan evaluation using dose–volume histogram (dvh) and spatial dose–volume histogram (zdvh)," *International Journal of Radiation Oncology* Biology* Physics*, vol. 43, no. 5, pp. 1143–1150, 1999.

[52] P. M. Medin and T. P. Boike, "Spinal cord tolerance in the age of spinal radiosurgery: lessons from preclinical studies," *International Journal of Radiation Oncology* Biology* Physics*, vol. 79, no. 5, pp. 1302–1309, 2011.

[53] A. Niroomand-Rad, C. R. Blackwell, B. M. Coursey, K. P. Gall, J. M. Galvin, W. L. McLaughlin, A. S. Meigooni, R. Nath, J. E. Rodgers, and C. G. Soares, "Radiochromic film dosimetry: recommendations of aapm radiation therapy committee task group 55," *Medical physics*, vol. 25, no. 11, pp. 2093–2115, 1998.

[54] P. A. Jursinic, "Angular dependence of dose sensitivity of surface diodes," *Medical physics*, vol. 36, no. 6Part1, pp. 2165–2171, 2009.

[55] M. Petasecca, S. Alhujaili, A. Aldosari, I. Fuduli, M. Newall, C. Porumb, M. Carolan, K. Nitschke, M. Lerch, J. Kalliopuska, *et al.*, "Angular independent silicon detector for dosimetry in external beam radiotherapy," *Medical physics*, vol. 42, no. 8, pp. 4708–4718, 2015.

[56] D. A. Low and J. F. Dempsey, "Evaluation of the gamma dose distribution comparison method," *Medical physics*, vol. 30, no. 9, pp. 2455–2464, 2003.

[57] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical physics*, vol. 25, no. 5, pp. 656–661, 1998.

[58] T. Fawcett, "An introduction to roc analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861–874, 2006.

[59] M. Greiner, D. Pfeiffer, and R. Smith, "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests," *Preventive veterinary medicine*, vol. 45, no. 1-2, pp. 23–41, 2000.

[60] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *Journal of biomedical informatics*, vol. 38, no. 5, pp. 404–415, 2005.

[61] P. A. Flach and S. Wu, "Repairing concavities in roc curves.," in *IJCAI*, pp. 702–707, 2005.

[62] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, vol. 3, no. 1, pp. 32–35, 1950.

[63] M. Carlone, C. Cruje, A. Rangel, R. McCabe, M. Nielsen, and M. MacPherson, "Roc analysis in patient specific quality assurance," *Medical physics*, vol. 40, no. 4, 2013.

[64] G. Yan, C. Liu, T. A. Simon, L.-C. Peng, C. Fox, and J. G. Li, "On the sensitivity of patient-specific imrt qa to mlc positioning errors," *Journal of applied clinical medical physics*, vol. 10, no. 1, pp. 120–128, 2009.

[65] J.-i. Kim, S.-Y. Park, H. J. Kim, J. H. Kim, S.-J. Ye, and J. M. Park, "The sensitivity of gamma-index method to the positioning errors of high-definition mlc in patient-specific vmat qa for sbrt," *Radiation Oncology*, vol. 9, no. 1, p. 167, 2014.

[66] L. Vieillevigne, J. Molinier, T. Brun, and R. Ferrand, "Gamma index comparison of three vmat qa systems and evaluation of their sensitivity to delivery errors," *Physica Medica*, vol. 31, no. 7, pp. 720–725, 2015.

[67] B. Liang, B. Liu, F. Zhou, F.-f. Yin, and Q. Wu, "Comparisons of volumetric modulated arc therapy (vmat) quality assurance (qa) systems: sensitivity analysis to machine errors," *Radiation Oncology*, vol. 11, no. 1, p. 146, 2016.

[68] W. Liu, X. Zhang, Y. Li, and R. Mohan, "Robust optimization of intensity modulated proton therapy," *Medical physics*, vol. 39, no. 2, pp. 1079–1091, 2012.

[69] G. Van Rossum and F. L. Drake Jr, *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[70] P. Mildenberger, M. Eichelberg, and E. Martin, "Introduction to the dicom standard," *European radiology*, vol. 12, no. 4, pp. 920–927, 2002.

[71] J. Y. Chang, A. Bezjak, F. Mornex, I. A. R. T. Committee, *et al.*, "Stereotactic ablative radiotherapy for centrally located early stage non–small-cell lung cancer: what we have learned," *Journal of Thoracic Oncology*, vol. 10, no. 4, pp. 577–585, 2015.

[72] G. Hanna, L. Murray, R. Patel, S. Jain, K. Aitken, K. Franks, N. Van As, A. Tree, P. Hatfield, S. Harrow, *et al.*, "Uk consensus on normal tissue dose constraints for stereotactic radiotherapy," *Clinical Oncology*, vol. 30, no. 1, pp. 5–14, 2018.

[73] S. N. Corporation, "Snc patient manual, v2.6.3, revision i.," Sep 2012.

[74] V. P. Keeling, S. Ahmad, and H. Jin, "A comprehensive comparison study of three different planar imrt qa techniques using mapcheck 2," *Journal of applied clinical medical physics*, vol. 14, no. 6, pp. 222–233, 2013.

[75] P. A. Jursinic, R. Sharma, and J. Reuter, "Mapcheck used for rotational imrt measurements: step-and-shoot, tomotherapy, rapidarc," *Medical physics*, vol. 37, no. 6Part1, pp. 2837–2846, 2010.

[76] J. Zhang, "Su-gg-t-247: A new method to compensate angular dependency of mapcheck device in intensity modulated arc therapy," *Medical Physics*, vol. 37, no. 6Part19, pp. 3242–3242, 2010.

[77] A. TG40, "Comprehensive qa for radiation oncology," *Medical Physics*, vol. 214, pp. 581–618, 1994.

[78] H. Jin, F. B. Jesseph, and S. Ahmad, "A comparison study of volumetric modulated arc therapy quality assurances using portal dosimetry and mapcheck 2," *Progress in Medical Physics*, vol. 25, no. 2, pp. 65–71, 2014.

[79] I. ISO and B. OIML, "Guide to the expression of uncertainty in measurement," *Geneva, Switzerland*, 1995.

[80] D. Létourneau, M. Gulam, D. Yan, M. Oldham, and J. W. Wong, "Evaluation of a 2d diode array for imrt quality assurance," *Radiotherapy and oncology*, vol. 70, no. 2, pp. 199–206, 2004.

[81] T. Sing, O. Sander, N. Beerenwinkel, and T. Lengauer, "Rocr: visualizing classifier performance in r," *Bioinformatics*, vol. 21, no. 20, pp. 3940–3941, 2005.

[82] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[83] P. S. Basran and M. K. Woo, "An analysis of tolerance levels in imrt quality assurance procedures," *Medical physics*, vol. 35, no. 6Part1, pp. 2300–2307, 2008.

[84] B. Mijnheer and D. Georg, "Estro booklet 9," *Guidelines for the verification of IMRT*, 2008.

[85] A. Fredh, J. B. Scherman, L. S. Fog, and P. Munck af Rosenschöld, "Patient qa systems for rotational radiation therapy: a comparative experimental study with intentional errors," *Medical physics*, vol. 40, no. 3, p. 031716, 2013.

[86] A. K. Templeton, J. C. Chu, and J. V. Turian, "The sensitivity of arccheck-based gamma analysis to manufactured errors in helical tomotherapy radiation delivery," *Journal of applied clinical medical physics*, vol. 16, no. 1, pp. 32–39, 2015.

[87] B. Mzenda, K. V. Mugabe, R. Sims, G. Godwin, and D. Loria, "Modeling and dosimetric performance evaluation of the raystation treatment planning system," *Journal of applied clinical medical physics*, vol. 15, no. 5, pp. 29–46, 2014.

[88] J.-Y. Park, S. Kim, H.-J. Park, J.-W. Lee, Y.-S. Kim, and T.-S. Suh, "Optimal set of grid size and angular increment for practical dose calculation using the dynamic conformal arc technique: a systematic evaluation of the dosimetric effects in lung stereotactic body radiation therapy," *Radiation Oncology*, vol. 9, no. 1, p. 5, 2014.

[89] H. Chung, H. Jin, J. Palta, T.-S. Suh, and S. Kim, "Dose variations with varying calculation grid size in head and neck imrt," *Physics in Medicine & Biology*, vol. 51, no. 19, p. 4841, 2006.

[90] G. M. Videtic, C. Hu, A. K. Singh, J. Y. Chang, W. Parker, K. R. Olivier, S. E. Schild, R. Komaki, J. J. Urbanic, and H. Choy, "A randomized phase 2 study comparing 2 stereotactic body radiation therapy schedules for medically inoperable patients with stage i peripheral non-small cell lung cancer: Nrg oncology rtog 0915 (ncctg n0927)," *International Journal of Radiation Oncology* Biology* Physics*, vol. 93, no. 4, pp. 757–764, 2015.

[91] T. Kron, B. Chesson, N. Hardcastle, M. Crain, N. Clements, M. Burns, and D. Ball, "Credentialing of radiotherapy centres in australasia for trog 09.02 (chisel), a phase iii clinical trial on stereotactic ablative body radiotherapy of early stage lung cancer," *The British journal of radiology*, vol. 91, no. 1085, p. 20170737, 2018.

[92] S. Siva, T. Kron, M. Bressel, M. Haas, T. Mai, S. Vinod, G. Sasso, W. Wong, H. Le, T. Eade, *et al.*, "A randomised phase ii trial of stereotactic ablative fractionated radiotherapy versus radiosurgery for oligometastatic neoplasia to the lung (trog 13.01 safron ii)," *BMC cancer*, vol. 16, no. 1, p. 183, 2016.

[93] D. Comsa, E. Barnett, D. Moseley, and M. van Prooijen, "Su-e-t-445: Use of mapphan for patient specific vmat plan verification," *Medical Physics*, vol. 38, no. 6Part17, pp. 3591–3591, 2011.

[94] S. De Boer and H. Malhotra, "Su-ff-t-204: The determination of optimal hounsfield unit assignment of materials in a 2 dimensional diode array used in the calculation of rotational therapy verification plans," *Medical Physics*, vol. 36, no. 6Part12, pp. 2567–2567, 2009.

[95] T. Depuydt, A. Van Esch, and D. P. Huyskens, "A quantitative evaluation of imrt dose distributions: refinement and clinical assessment of the gamma evaluation," *Radiotherapy and oncology*, vol. 62, no. 3, pp. 309–319, 2002.

[96] E. M. McKenzie, P. A. Balter, F. C. Stingo, J. Jones, D. S. Followill, and S. F. Kry, "Toward optimizing patient-specific imrt qa techniques in the accurate detection of dosimetrically acceptable and unacceptable patient plans," *Medical physics*, vol. 41, no. 12, 2014.

[97] L. Dong, J. Antolak, M. Salehpour, K. Forster, L. ONeill, R. Kendall, and I. Rosen, "Patient-specific point dose measurement for imrt monitor unit verification," *International Journal of Radiation Oncology Biology Physics*, vol. 56, no. 3, pp. 867–877, 2003.

[98] J. Stroom, S. Vieira, D. Mateus, C. Greco, A. Fogliata, G. Nicolini, A. Clivio, E. Vanetti, and L. Cozzi, "On the robustness of vmat-sabr treatment plans against isocentre positioning uncertainties," *Radiation Oncology*, vol. 9, no. 1, p. 196, 2014.

[99] D. J. Hand, "Measuring classifier performance: a coherent alternative to the area under the roc curve," *Machine learning*, vol. 77, no. 1, pp. 103–123, 2009.

[100] B. M. Rauschenbach, L. Mackowiak, and H. K. Malhotra, "A dosimetric comparison of three-dimensional conformal radiotherapy, volumetric-modulated arc therapy, and dynamic conformal arc therapy in the treatment of non-small cell lung cancer using stereotactic body radiotherapy," *Journal of applied clinical medical physics*, vol. 15, no. 5, pp. 147–161, 2014.

[101] M. K. Tyler, "Quantification of interplay and gradient effects for lung stereotactic ablative radiotherapy (sabr) treatments," *Journal of applied clinical medical physics*, vol. 17, no. 1, pp. 158–166, 2016.

[102] K. Nithiyanantham, G. K. Mani, V. Subramani, L. Mueller, K. K. Palaniappan, and T. Kataria, "Analysis of direct clinical consequences of mlc positional errors in volumetric-modulated arc therapy using 3d dosimetry system," *Journal of applied clinical medical physics*, vol. 16, no. 5, pp. 296–305, 2015.

[103] A. Agnew, C. Agnew, M. Grattan, A. Hounsell, and C. McGarry, "Monitoring daily mlc positional errors using trajectory log files and epid measurements for imrt and vmat deliveries," *Physics in Medicine & Biology*, vol. 59, no. 9, p. N49, 2014.

[104] X. Wu, J. Kalliopuska, S. Eränen, and T. Virolainen, "Recent advances in processing and characterization of edgeless detectors," *Journal of Instrumentation*, vol. 7, no. 02, p. C02001, 2012.

[105] A. B. Rosenfeld, "Mosfet dosimetry on modern radiation oncology modalities," *Radiation protection dosimetry*, vol. 101, no. 1-4, pp. 393–398, 2002.